

Testgütekriterien von Erhebungsinstrumenten zur Erforschung egozentrierter sozialer Netzwerke

LV Handlungsleitende Theorien der Methoden der Sozialarbeit

bei FH-Prof. Dr. Peter Pantucek

Seminararbeit von Christoph Pammer, DSA, MPH

FH Joanneum |SAG08

Graz, April 2009

Inhaltsverzeichnis

1. Einleitung	2
2. Exzerpt eines Systematic Reviews zum Thema	2
3. Exkurs: Testgütekriterien: Validität	7
3.1 Reliabilität.....	7
3.2 Änderungssensitivität	8
4. Die Testgüte und Datenqualität zur Bestimmung egozentrierter sozialer Netzwerke ...	10
5. Fazit	13

1. Einleitung

Auch aus dem Blickwinkel der Gesundheitswissenschaften und von Public Health ist die empirische Grundlagenforschung und theoretische Auseinandersetzung mit den sozialen Netzwerken von großer Bedeutung. Eine der bedeutsamsten Fragen richtet sich auf die Entstehungstheorien für Gesundheit: Wie entsteht soziale Gesundheit? Da Gesundheit überwiegend sozial produziert wird, geht diese Frage weit über den Erkenntnishorizont einzelner wissenschaftlicher Disziplinen hinaus. Anders als in den Sozialwissenschaften liegt der Fokus innerhalb der theoretischen Konzepte noch stärker auf „sozialer Unterstützung“ als Output und Outcome infolge sozialer Beziehungen. Dass soziale Unterstützung eine wichtige präventive und kurative Gesundheitsressource darstellt und unterschiedlich in der Bevölkerung verteilt ist, ist mithin bekannt. Die sozial bedingte gesundheitliche Ungleichheit darf weder vom Sozial- noch vom Gesundheitssystem außer Acht gelassen werden.

Als Synthese beider Perspektiven schlage ich deshalb aus verschiedenen Gründen vor, die soziale Netzwerkforschung als Analyseinstrument der Gesundheitsentstehung und Krankheitsverhütung und -behandlung verstärkt heranzuziehen. Zum einen sind die in gesundheitswissenschaftlichen Studien herangezogenen Parameter zur Einschätzung sozialer Netzwerke unterentwickelt, weil soziale Netzwerke kaum im Fokus der Betrachtung standen. Da aber beispielsweise auf Populationsebene bekannt ist, dass die Lebenserwartung und der Zuwachs an Lebenserwartung in erster Linie vom Ausmaß gesundheitlicher Ungleichheit abhängig ist und auch eindeutig mit dem Sozialkapital in einer Gesellschaft korreliert, hat sich der Fokus der Gesundheitsförderungsforschung längst auch auf die Erforschung der Lebensbedingungen des Menschen in seiner sozialen und strukturellen Umwelt erweitert.

Diese Seminararbeit widmet sich der Frage, in wieweit die Instrumente zur systematischen Erforschung sozialer Netzwerke statistisch und theoretisch belastbarer sind, weil Klarheit der Konzepte und korrektes Messen eine Voraussetzung für die Weiterentwicklung der Forschung an der Schnittmenge von Gesundheit und Sozialem voraussetzt. Liegt ein zuverlässiges Messinstrument vor, so können mit diesem auch zuverlässige Ergebnisse erzielt werden.

2. Exzerpt eines Systematic Reviews zum Thema

Christof Wolf erarbeitet in seinem Überblicksartikel „**Egozentrierte Netzwerke. Erhebungsverfahren und Datenqualität**“ die Geschichte und Anwendung der systematischen empirischen Erhebung egozentrierter Netzwerke. Als Folge des bei Ulrich Beck (Beck, 1986, 208ff) und anderen beschriebenen Individualisierungsschubs in der Gesellschaft geriet das Individuum in den Mittelpunkt der soziologisch-theoretischen Betrachtung: „Statt weiterhin `die Familie`, `die Klassenzugehörigkeit` oder `die Nachbarschaft` zu untersuchen, lenkte die Individualisierung die Aufmerksamkeit der Forschung immer stärker auf das Netzwerk faktisch bestehender Beziehungen.“ (Wolf, 2006, S. 245). Als eine weitere Entwicklung in der soziologischen Theorie, die dem Netzwerkkonzept Auftrieb gab und weiterhin gibt, war die Entdeckung sozialer Beziehungen als Ressource, als soziales Kapital.

Der Ansatz egozentrierter Netzwerkforschung stellt das Individuum - die sozialen Beziehungen einer Fokuspersion - ins Zentrum der Betrachtungsweise. Von ihr/ihm ausgehend werden die Beziehungen zu anderen Personen definiert. (Die Alternative wäre das Konzept des Gesamtnetzwerks, worunter die Menge von sozialen Beziehungen aller Personen innerhalb von klar abgrenzbaren Gruppen zu verstehen wäre). Die Erhebung egozentrierter Netzwerke besteht aus drei Teilen:

- einer oder mehrerer Fragen nach Personen, mit denen die befragte Person in Beziehung steht (Namensgeneratoren)
- Fragen nach den Eigenschaften der Netzwerkpersonen (Alter) und zur Beziehung zwischen Ego und Alter (Namensinterpretatoren)
- Fragen nach der Beziehung der Netzwerkgeneratoren untereinander

Nach den ersten Gehversuchen der egozentrierten Netzwerkerhebung durch Laumann (1963, zit. n. Wolf, 2006, S. 246) entwickelten McCallister und Fischer (1978, z. n. Wolf, 2006, S. 246) ein Instrument, welches sie in der North California Community Study einsetzten. Das Instrument zielte darauf ab, möglichst viele Netzwerkpersonen aus möglichst unterschiedlichen sozialen Kontexten zu erfassen und wurde in mehreren Studien bis in die Mitte der 1990er Jahre eingesetzt. Die Items waren so formuliert, dass wir heute sagen würden, es handelte sich um eine Abfrage der sozialen Unterstützung, welche aus definierten Netzwerkkontakten gezogen werden konnte. Konzeptionell führen solche Verfahren dazu, relativ große Netzwerke zu erzeugen, die sich bspw. nach unterschiedlichen Interaktionsformen oder Beziehungsarten differenzieren lassen. Die Erhebungszeit des Instruments von McCallister und Fischer beträgt jedoch 20 Minuten, weswegen es sich nur für Studien eignet, deren Schwerpunkt auf der Erhebung sozialer Netzwerke liegt und die ausreichend Mittel zur Erhebung zur Verfügung stellen können. Für den Fall, dass die sozialen Netzwerke nicht im Mittelpunkt des Interesses stehen, hat Burt (1984, z. n. Wolf 2006, S. 247) einen Namensgenerator vorgeschlagen, der zuerst im General Social Survey 1985 eingesetzt wurde. Dieser besteht aus einer einzigen Frage: „Hin und wieder besprechen die meisten Leute wichtige Angelegenheiten mit anderen. Wenn Sie an die letzten sechs Monate zurückdenken: Mit wem haben Sie über Dinge gesprochen, die Ihnen wichtig waren?“. Wir Burt anhand der von Fischer aufwändiger erhobenen Daten zeigen konnte, handelt es sich bei den so erhobenen Personen um eine Teilgruppe des egozentrierten Netzwerks, die durch relativ nahe Beziehungen (intimacy) gekennzeichnet sind, mit einem höheren Anteil von Freunden und Verwandten. Die bei Wolf gezeigte Tabelle stellt diese Untersuchungen wie folgt gegenüber:

Tabelle 1: Die Namensgeneratoren von McCallister, Fischer und Burt im Vergleich

	McCallister und Fischer					Burt		
	NCCS ^a	ZUMA ^b	JV ^c	MC ^c	Haifa ^d	GSS ^e	ZUMA ^b	WS87 ^f
Anzahl Befragte	1.050	239	98	99	262	1.531	240	1.311
Ø Netzwerkgröße	18,5	7,8	21,8	10,1	14,0	2,9	2,6	2,5
Prozent Verwandte	42	39	40	46	44	55	49	60

a North California Community Study (Fischer 1982).

b ZUMA-Methodenexperiment „Egozentrierte Netzwerke“ (Pfenning und Pfenning 1987).

c JV: Jacksonville, Florida, und MC: Mexico-City; Methodenexperiment „Measuring Personal Social Networks“ (Bernard et al. 1990).

d Haifa-Studie (Fischer und Shavit 1995).

e General Social Survey 1985; im GSS konnten bis zu sechs Personen genannt werden; um den Vergleich mit den beiden anderen Studien zu gewährleisten, wurde diese Zahl hier auf maximal fünf festgelegt.

f Wahlstudie 1987 (Pappi und Melbeck 1988: 233).

Erwartungsgemäß sind die mit dem Generator nach McCallister und Fischer erhobenen Netzwerke deutlich größer als die nach der Burt-Methode erfassten. Darüberhinaus zeigt sich ein deutlich höherer Anteil Verwandter bei der Methode nach Burt als bei McCallister und Fischer, bei denen der Anteil immerhin auch noch mehr als 40 % beträgt.

Die frühen Instrumente versuchen, positiv bewertete Beziehungen zu erheben. Der Versuch, negativ bewertete Beziehungen zu erheben, findet sich hingegen kaum. Straits (2000, z.n. Wolf 2006, S. 248) hat daher untersucht, ob der Burt-Generator mit dem Zusatz „... these may include people that sometimes make you angry or upset“ zu anderen Ergebnissen führt, was anscheinend nicht der Fall war. Wolf entwickelte dieses Konzept deshalb dahingehend weiter, dass im zweiten Teil des Netzwerkinstruments für jede Beziehung zu erheben wäre: „Beziehungen zu anderen Menschen können hilfreich, aber manchmal auch belastend sein. Wie häufig ist Ihre Beziehung zu ... (Name) für Sie belastend?“. In der ersten Studie zeigte sich, dass eine größere Belastung durch das soziale Netzwerk mit einer geringeren psychischen Gesundheit korreliert. (Wolf, 2003 ZITIEREN).

Ende der 1990er Jahre entwickelte Burt in Zusammenarbeit mit anderen und in Anlehnung an McCallister und Fischer Netzwerkinstrumente mit neun bzw. zehn Generatorfragen für spezifische Zielgruppen (hier: Manager/innen). Van der Poel (1995, z. n. Wolf, 2006, S. 249) war stärker an der sozialen Unterstützung interessiert und entwarf einen zehn Fragen umfassenden Namensgenerator, bei dem hypothetische Fragen verwendet werden. Eine weitere Methode wurde von Wolf und Kecskes (1996) entwickelt und seither in vier weiteren Studien modifiziert eingesetzt. Die Namensgeneratoren beziehen sich dabei auf konkrete soziale Interaktionen. Die Generierung der Items wurde von bestehenden empirischen Analysen und theoretischen Überlegungen zur Bedeutung instrumenteller und geselliger Kontakte geleitet. Dies führte dazu, dass die Interviewer angeleitet wurden, keine Personen im beruflichen Umfeld oder Ärzt/innen, Priester oder Pfleger/innen als Netzwerkknoten zuzulassen. Im Gegensatz zu früheren Erhebungen wurde der Zeitraum, auf der sich die Fragen beziehen, auf die letzten 14 Tage begrenzt, um die Zuverlässigkeit der Information zu erhöhen.

Die bisher genannten Instrumente hatten zum Ziel, einen mehr oder weniger großen Ausschnitt der sozialen Beziehungen der Egos zu erheben, der durch regelmäßige Interaktion gekennzeichnet ist, wobei die Art der Interaktion durch die Formulierung des Namensgenerators vorgegeben ist. Wolf bezeichnet diese Formen als aufwändig, aber sehr flexibel. Daneben gibt es jedoch auch andere Ansätze, die seltener eingesetzt werden, bspw. von McCarthy et al. (1997, z. n. Wolf 2006, S. 251), der versucht, valide Punktschätzer für den Umfang und teilweise auch für die Zusammensetzung egozentrierter Netzwerke (allerdings in) Gesamtnetzwerken zu erhalten.

Für die Soziale Arbeit bedeutsamere Ansätze (Anm. d. Verf.) sind an den wichtigen Bezugspersonen und der emotionalen Verbundenheit des Egos zu diesen interessiert. Kahn und Antonucci (1980, z. n. Wolf, 2006, S. 250) entwickelten eine Methode, bei der Befragten ein Zielscheibendiagramm mit drei konzentrischen Kreisen vorgegeben wird. Die sehr eng verbundenen Personen sollen in den ersten Kreis, die eng verbundenen Personen in den zweiten Kreis und weitere wichtige Personen, mit denen jedoch keine enge Verbindung besteht, in den dritten Kreis positioniert werden. Mit dieser Methode werden wie in den anderen Verfahren auf der Basis von Namensgeneratoren Verwandte, zu denen eine engere emotionale Beziehung besteht, erfasst, aber als Erweiterung auch emotional bedeutsame Beziehungen zu entfernteren Verwandten und Nicht-Verwandten erhoben (Wagner et al., 1996, z. n. Wolf, 2006, S. 254). Wenn man nicht an der Zusammensetzung und konkreten Struktur (Namen) interessiert ist, sondern lediglich an der sozialen Reichweite, eignet sich der Positionsgenerator von Lin und Dumin (1986, z. n. Wolf, 2006, S. 252). Bei dieser Methode wird den Befragten eine Liste mit sozial bedeutsamen Positionen vorgelegt, z.B. Berufe, Ämter oder auch Wohnorte, und sie sollen angeben, ob sie Personen kennen, die die entsprechenden Positionen innehaben. Der Vorteil erstreckt sich auf einen vergleichsweise geringen Erhebungsaufwand. Dieser Positionsgenerator dient dazu, denjenigen Ausschnitt des sozialen Kapitals einer Person zu erheben, der vor allem für instrumentelle Ziele, z. B. dem Finden eines neuen Arbeitsplatzes, genutzt werden kann. Soll dagegen ein umfassendes Maß für die Sozialkapitalausstattung einer Person erhoben werden, bietet sich die Verwendung des von Snijders et al. (2004a, z. n. Wolf, 2006, S. 253) vorgeschlagenen Ressourcengenerators an. Dieses Instrument besteht aus 35 Items („Do you know anybody who... als Impuls und Auflistung von 35 sozialen Ressourcen anhand von Beispielen; Beantwortung „ja/nein“).

Wolf weist darauf hin, dass die Auswahl von Namensgeneratoren, die dazu führt, dass gewisse Bereiche des sozialen Netzwerks fokussiert werden, von der Forschungsfrage abhängig ist. Ein praktischer Hinweis ergibt sich als Festlegung eines Mindestalters für Alteri, da nachfolgende Items bei besonders jungen Personen nicht vervollständigt werden konnten. Ein zweites praktisches Problem bei der Erhebung sozialer Netzwerke in Studien ergibt sich bzgl. Personen ohne soziales Netzwerk und deren Unterscheidung von Antwortverweigerern. Das Instrument sollte explizit beide Fälle vorsehen.

Wie weiter oben erwähnt werden die Merkmale der Netzwerkpersonen über sog. Namensinterpretatoren erhoben. Dies ist bspw. inhaltlich relevant, weil die Ähnlichkeit von Netzwerkpersonen in Bezug auf sozio-ökonomische Merkmale oft als Gradmesser für die Offenheit von Gesellschaften angesehen wird. Im Zusammenhang mit Namensinterpretatoren muss unterschieden werden,

- ob die Merkmale für alle oder nur einen Teil der genannten Netzwerkpersonen unterschieden werden sollen (abhängig von Fragestellung), da damit Selektionseffekte (stärkere Beziehungen) verbunden sein können. Aufgrund der Begrenztheit der Befragungsressourcen (Zeit für Ausfüllen des Fragebogens, in Interviews, Geld etc.) steigt die Dringlichkeit für so eine Auswahl mit der Größe des Netzwerks, initiiert durch die Namensgeneratoren. ZUMA führte dazu ein Methodenexperiment durch, bei dem die Auswahl der wichtigsten 5 Personen den Befragten überlassen wurde. Eine andere Strategie besteht darin, aus der Menge der generierten Personen eine Zufallsauswahl zu treffen, mit dem Vorteil, systematischer Verzerrungen durch Positivselektion der Personen vorzubeugen und Aussagen über das gesamte erhobene Netzwerk zuzulassen. Solche Auswahlverfahren, die der/m Interviewer/in obliegen, stellen aber ihrerseits eine Fehlerquelle dar, die sich auf die Testgüte auswirkt (Interrater Reliabilität).
- Die zweite Frage im Zusammenhang mit den Angaben zu Netzwerkpersonen ist, ob diese Angaben personen- oder merkmalsbezogen erhoben werden sollen. Bei der personenbezogenen Methode werden die Merkmale einer Person nach der anderen erfasst, bei der merkmalsbezogenen wird ein Merkmal nach dem anderen für alle Personen erfasst. Ein neueres Methodenexperiment zeigt, dass die personenbezogene Vorgehensweise zuverlässigere Daten liefert (Kogovsek et al. 2002, z. n. Wolf 2006, S. 257).

Ein zentrales Strukturmerkmal egozentrierter Netzwerke ist deren Dichte. Zur Berechnung dieses Merkmals werden Angaben über die Beziehungen der Alteri untereinander benötigt. Das Erhebungsproblem adressiert hier die mögliche Anzahl von Beziehungen der Alteri zueinander, die bei 5 Personen 10, bei 10 Personen aber bereits Nachfragen zu 45 Beziehungen zur Folge hat. Bei 20 Beziehungen wären es 190 potentielle Beziehungen. Deshalb beschränkt man sich auf eine Auswahl von Alteri zur Abfrage der Beziehungen untereinander, die nach Möglichkeit (oder Fragestellung) zufällig ausgewählt werden sollten. Die Items zielen inhaltlich typischerweise darauf ab, ob sie einander kennen bzw. auf die Beziehungsstärke. Auf Fragen wie „Stehen diese sich einander sehr nahe.. nahe.. etc.“ folgen 3 bis 5stufige Skalierungen, wobei unklar bleibt, welchen Einfluss die Skalierung auf die Qualität der Dichtemessung nimmt.

3. Exkurs: Testgütekriterien: Validität

Die Biostatistik unterscheidet eine Vielzahl von Testverfahren. Die Validität von Messungen sozialwissenschaftlicher Indikatoren ist definiert als die Fähigkeit eines Tests, z.B. zwischen Personen mit bestimmten Netzwerkeigenschaften und Personen, bei denen diese Merkmale nicht vorliegen, zu unterscheiden. Die qualitative Beurteilung der Testverfahren unterliegt einer Reihe von systematisch anzuwendenden Testgütekriterien, die in diesem Abschnitt erläutert werden.

Die Validität beschreibt die Gültigkeit von Messungen von Eigenschaften egozentrierter sozialer Netzwerke. Da die genaue Verteilung verschiedener Netzwerkeigenschaften in Studienpopulationen nicht bekannt ist, wird bei der statistischen Beurteilung der Validität auf eine Reihe von in sich zusammenhängenden Hypothesen der Psychometrie Rücksicht genommen.

Ein erstes dieser Kriterien ist, ob der Inhalt des zu beurteilenden Messinstruments augenscheinlich dazu geeignet ist, das Richtige zu messen (*face validity, content validity*), was i. d. R. von einzelnen oder mehreren Expert/innen im Konsens festgestellt wird. Bei der Augenscheinvalidität handelt es sich um ein streitbares Testgütekriterium, da Expertenmeinungen sich auch erheblich unterscheiden können, wenn es am empirischen Methodenverständnis mangelt. Die Psychometrie stützt ihre Aussagen deshalb auf empirisch gesicherte Information.

a. Reliabilität

Die Reliabilität ist ein Testgütekriterium, welches das Ausmaß an sowohl systematischen als auch zufälligen Fehlern beschreibt, das jeder Messung inhärent ist. Die Reliabilität beschreibt die Fähigkeit eines Tests, eine Testpopulation richtig zu kategorisieren und diese Eigenschaft unter veränderten Bedingungen beizubehalten.

Die innere Konsistenz des Instruments kann anhand von Messungen zu einem einzelnen bestimmten Zeitpunkt beurteilt werden. Man vergleicht, wie stark die Ergebnisse einzelner Subskalen des Instruments zueinander in Beziehung stehen und berechnet statistische Maßzahlen (Cronbachs alpha, Kuder-Richardson usw.). In der Regel werden (verschiedene Arten) von Korrelations-Koeffizienten auf der Basis erhobener Daten berechnen, die ein Maß für die Gleichgerichtetheit zweier Eigenschaften sind. Ein positiver Korrelations-Koeffizient bedeutet, dass ein gleichsinniger Zusammenhang in einem Vergleich der Stichproben besteht. Messwerte über 0 (max. 1) bezeichnen einen Gleichsinnigen Zusammenhang. Bei den in den besprochenen Arbeiten angewendeten statistischen Methoden und Möglichkeiten (Test-Retest-Reliabilität; Interrater-Rater Reliabilität) kann allgemein von einem Korrelations-Koeffizient $r > 0,75$ auf ausgezeichneter Testgüte geschlossen werden, $r < 0,4$ lässt auf mangelhafte Testgüte schließen.

Die Stabilität von Messungen kann zu unterschiedlichen Zeitpunkten (Test-Retest-Reliabilität) bzw. von unterschiedlichen Personen (Inter-Observer-Reliabilität synonym Interrater Reliabilität) bzw. von denselben Personen aber zu unterschiedlichen Zeitpunkten gemessen (Intra-Observer-Reliabilität synonym Intrarater Reliabilität) stark variieren. Die Ergebnisse, die Klient/innen sozialer Arbeit oder

Studienpopulationen auf den Instrumenten erzielen, variieren aufgrund einer Vielzahl von Faktoren, etwa der Anzahl der zu Messungen eingesetzten Items, der Verteilung sozio-demographischer Charakteristika in der Studienpopulation, systematischer Unterschiede der Klient/innenkollektive und auch, wenn Tests zu unterschiedlichen Zeitpunkten von unterschiedlichen Personen durchgeführt werden.

Während die Reliabilität darüber informiert, ob die Testergebnisse reproduzierbar sind, sagt sie nichts darüber aus, was gemessen wurde. In der Regel werden die Testverfahren deshalb mit den bereits etablierten Verfahren verglichen (konkurrente Validität, konvergente Validität). Dabei gelten die bestehenden, etablierten Skalen, mit denen ein neu entwickeltes Instrument verglichen wird als „Gold Standard“.

b. Änderungssensitivität

Die Veränderungssensitivität ist die Fähigkeit eines Instruments, praktisch relevante Veränderungen mit einiger Zuverlässigkeit zu messen. Aus dieser Perspektive geraten Reliabilität und Validität nicht in den Hintergrund, sondern sind viel mehr Voraussetzungen für das Messen von Veränderungen in egozentrierten sozialen Netzwerken. Aus messtheoretischer Perspektive ist so möglich, die Ergebnisse der Behandlung abzubilden und damit die Effektivität der sozialarbeiterischen Interventionen zu untersuchen.

Die Ziele der Veränderungsmessung selbst sind:ⁱ

- die inter-individuellen Unterschiede im Ausmaß der Veränderung zu bestimmen und Subgruppen zu identifizieren, in denen die Interventionen effektiver sind als in anderen;
- die Korrelationen der Veränderung zu bestimmen, um jene Faktoren zu identifizieren, die mit den Veränderungen zusammenhängen.

Wenn sich bei Messungen egozentrierter sozialer Netzwerke zu unterschiedlichen Zeitpunkten herausstellt, dass manche Klient/innen größere Verbesserungen zeigen als andere, ist es schwierig, den Effekt der Intervention selbst zu bestimmen. In experimentellen Studien wird durch Randomisierung eine zufällige Verteilung dieser Einflussfaktoren angenommen, um prognostische Faktoren der Wirksamkeit einer Intervention bestimmen zu können. In der Front Line Sozialarbeit hingegen, in der die „Wirksamkeit unter realen Bedingungen“ zu eruieren ist, ist es wünschenswert, zwischen der Wirksamkeit verschiedener Interventionsstrategien in Subgruppen von Patient/innen unterscheiden zu können, um Aussagen über die Durchführungsqualität (Performanz) der Leistungen treffen zu können.

Das methodische Vorgehen zur Bestimmung der Änderungssensitivität kann zweierlei sein: Entweder man unterscheidet im Vorhinein zwei Gruppen von Klient/innen anhand der Ergebnisse einer Messung zum Zeitpunkt T1 voneinander – und zwar anhand der Wahrscheinlichkeit, mit der eine festzulegende Veränderung eintritt – und berechnet die Effektstärke nach Guyattⁱⁱ oder die „Receiver Operator Characteristic (ROC) Curve nach Deyo and Centor.ⁱⁱⁱ Nach der zweiten Methode werden Veränderungen innerhalb desselben Patient/innen-Kollektivs anhand der Effektstärke nach Kazis et al. beobachtet^{iv}, oder es wird nach statistisch signifikanten Veränderungen etwa anhand von t-Tests und anderen Methoden geprüft.

Die Änderungssensitivität eines Instruments oder Mess-Verfahrens ist in der Praxis von vielen Umweltbedingungen abhängig, die zu systematischen Fehlern in der Messung führen können. Würde man beispielsweise in Studien versuchen, die Klient/innen direkt nach den Verbesserungen durch die Intervention zu fragen, wie dies im praktischen Kontext Gang und Gäbe ist („Hat sich Ihr Befinden verbessert, seit Sie das letzte Mal da waren? Wie sehr?“), erhielte man ein schwer einschätzbares Ausmaß an verfälschten Antworten. Menschen neigen dazu, sich an den Ausgangszustand nur schwer erinnern zu können und machen deshalb Angaben, die stark mit der aktuellen Situation korrelieren. Eine Reihe von psychometrischen Studien^v ist der Frage nachgegangen, wie dieser Effekt die Ergebnisse verzerren kann. Etwa:

- Die Patient/innen unterschätzen den Ausgangszustand systematisch, was tendenziell zu einer Überschätzung des Interventionseffekts führt.
- Die fälschliche Einschätzung des Ausgangszustands korreliert mit dem Zustand zum Befragungstag (Cluster Bias)

Darüber hinaus sind unter den „realen“ Bedingungen der sozialarbeiterischen Intervention auch die spezifische Zusammenstellung des Klient/innen-Kollektivs, insbesondere die Verteilung von Alter und Schweregrad (sog. „Case mix“) als klassische Verzerrungsfaktoren zu berücksichtigen. Klient/innen könnten zudem aus weiteren Gründen (Befragungssituation, soziale Erwünschtheit) dazu neigen, einen insgesamt besseren oder auch schlechteren Zustand als den tatsächlichen anzugeben.

Die Aufgabe für die Testkonstruktion und -administration besteht darin, diese und weitere systematische Verzerrungen weitestgehend zu reduzieren. Wenn der Status eines egozentrierten sozialen Netzwerks bspw. von Sozialarbeiter/innen und nicht von Klient/innen selbst und/oder von Angehörigen beurteilt wird, kommt es in der Regel auch zu unterschiedlichen Verzerrungen. Aus der Medizin ist bekannt, dass Ärzte/innen den Gesundheitszustand zu Beginn einer Behandlung unterschätzen und bei der Entlassung überschätzen, wenn man ihre Bewertungen mit den Bewertungen anderer in der Behandlung tätigen Gesundheitsberufen, den Selbstbewertungen der Patient/innen oder Bewertungen durch Angehörige vergleicht.

4. Die Testgüte und Datenqualität zur Bestimmung egozentrierter sozialer Netzwerke

Wolf (Wolf, 2006, S. 258-268) hat die jüngste Übersichtsarbeit zu Erhebungsverfahren und Datenqualität egozentrierter sozialer Netzwerke verfasst. Darin stellt er vorerst als wichtige Frage heraus, wie stabil die Auskünfte, die Ego über eine mit ihm verbundene Netzwerkperson gibt, sind und ob die von Ego gemachten Angaben mit den tatsächlichen Eigenschaften Alters übereinstimmen. Als eine wichtige Quelle für solche Informationen gilt das ZUMA-Methodenforschungsprojekt „Egozentrierte Netzwerke in Massenumfragen“ (Hoffmeyer-Zlotnik 1987; Mohler und Pfenning 1987; Schneid 1987), in dessen Rahmen 720 Personen im Abstand von drei Wochen unter Verwendung verschiedener Namensgeneratoren befragt wurden.

Die Zuverlässigkeit der Netzwerkgeneratoren nach Fischer stellte sich als sehr unterschiedlich heraus. Die höchste Zuverlässigkeit wird bei den Generatoren „Hobbies besprechen“, „persönliche Dinge besprechen“ und „gemeinsame Aktivitäten wie bspw. Ausgehen“ erreicht. Die Reliabilität der Netzwerkgröße beträgt bei McCallister und Fischer (Test-Retest-Reliabilität) $r=0,77$ und liegt beim Burt-Generator lediglich bei $r=0,65$. Bien, Marbach und Neyer (1991, z. n. Wolf 2006, S. 259) berichten, dass die Stabilität von Namensgeneratoren bei aggregierten Daten sehr hoch ist, während auf individueller Ebene größere Unterschiede feststellbar sind. Nur etwa ein Drittel der Befragten nannte zu beiden Zeitpunkten exakt dieselbe Anzahl von Netzwerkpersonen. Die durchschnittliche Abweichung zwischen den zwei Erhebungszeitpunkten betrug 1,6 Personen, dennoch wurden drei Viertel aller Netzwerkpersonen zu beiden Erhebungszeitpunkten genannt. Insgesamt sei die Zuverlässigkeit von Namensgeneratoren also zufrieden stellend, wobei nicht übersehen werden kann, dass es noch zu wenige Studien gibt, die die Testgüte von Namensgeneratoren untersuchten.

Andere Studien fragen nach der Qualität der Angaben Egos über Alter und die Beziehung. Die Angabe von soziodemographischen Merkmalen (Geburtsjahr $r=0,98$; sozio-ökonomischer Status $r=0,89$; Religionszugehörigkeit $r=0,98$) können als gültig angesehen werden, wenn man die Angaben der Egos über Alters vergleicht. (Laumann, 1969, S. 60; z. n. Wolf, 2006, S. 259). Weniger gut waren Angaben zur Parteipräferenz der Alters ($r=0,73$). Pfenning zeigt in der folgenden Tabelle erstaunliche Zuverlässigkeit der Angaben der Ego über sich selbst (Spalte Ego über Ego) sowie über Alteri (Spalte Ego über Alter):

Tabelle 2: Reliabilitäten von Eigenangaben der Befragten und Angaben der Befragten über ihre Netzwerkpersonen im Vergleich (Test-Retest-Ergebnisse)

	Ego über Alter (N = 123)		Ego über Ego (N = 86)	
	% übereinstimmende Angaben	Reliabilität ^b	% übereinstimmende Angaben	Reliabilität ^b
<i>Demographische Angaben</i>				
Geschlecht	94	$\varphi \geq 0,88$	95	$\varphi \geq 0,90$
Geburtsjahr	68	$r \geq 0,96$	96	$r \geq 0,99$
Schulabschluss (7) ^a	77	$V \geq 0,76$	86	$V \geq 0,81$
Familienstand (9)	75	$V \geq 0,55$	98	$V \geq 0,97$
Haushaltsgröße	71	$r \geq 0,72$	96	$r \geq 0,97$
Konfession (6)	68	$V \geq 0,69$	83	$V \geq 0,68$
<i>Einstellungen</i>				
Religiosität (10)	29	$r \geq 0,34$	35	$r \geq 0,72$
Parteieneigung (2)	84	$\varphi \geq 0,67$	74	$\varphi \geq 0,46$
Parteiwahl (5)	85	$V \geq 0,83$	97	$V \geq 0,93$
Autoritarismus				
Disziplin (5)	42	$r = 0,55$	43	$r = 0,61$
Autorität (5)	35	$r = 0,56$	43	$r = 0,48$
<i>Erziehungsziele</i>				
Kritikfähigkeit (7)	41	$r = 0,29$	30	$r = 0,23$
Selbstvertrauen (7)	34	$r = 0,51$	45	$r = 0,39$
Fleiß (7)	34	$r = 0,57$	39	$r = 0,25$
Schulleistungen (7)	34	$r = 0,46$	49	$r = 0,47$

Für fast alle Merkmale gilt: Je höher die Reliabilität der Eigenangaben, desto höher die Reliabilität der Angaben der Befragten über Netzwerkpersonen.

Andere Studien betrachteten die Zuverlässigkeit der Angaben Egos über die Eigenschaften der sozialen Netzwerke. Hammer (1984, S 345 ff; z. n. Wolf, 2006, S. 260) konnte auf Daten zurückgreifen, die einen direkten Vergleich von Egos und Alters diesbezüglich erlauben. Sehr hohe Übereinstimmung erzielen dabei Angaben zur Kontakthäufigkeit, der Dauer der Beziehung, des Verwandtschaftsgrades und der Stärke der Beziehung (zwischen 79 und 100 %), weshalb Hammer die Qualität von Netzwerkdaten generell sehr hoch einschätzt. Wie Schenk et al. (1992, S. 102 ff) zeigen, variiert die Genauigkeit der Angaben Egos über Alteri NICHT mit der Dauer, der Kontakthäufigkeit, der Intensität oder dem Kontext der Beziehung. Die Angaben für ein „Kernnetzwerk“ sind dabei zuverlässiger als für ein weiter gefasstes Netzwerk.

. Unterschiedliche Datenqualität aufgrund von Erhebungsmethoden

Egozentrierte Netzwerkdaten wurden bisher in den allermeisten Fällen in persönlich-mündlichen Interviews erhoben, so auch in den bisher zitierten Studien. Kogovsek et al (2002) untersuchten, ob die Datenqualität von der Erhebungsmethode abhängt anhand der Merkmale Häufigkeit des Kontakts, subjektive Nähe, Wichtigkeit der Beziehung und Häufigkeit, mit der sich Ego über Alter ärgert. Diese Merkmale wurden im Rahmen einer Face-to-Face-Befragung und mittels telefonischer Befragung, in beiden Fällen computerunterstützt, erhoben. Die Telefoninterviews lieferten in diesem Methodenexperiment die besseren Validitätswerte, wofür die Autoren die höhere Anonymität bei telefonischen Befragungen verantwortlich machen. Auch bezüglich aller Reliabilitätswerte schneidet die telefonische Befragung deutlich besser ab. Das heißt, dass die relativ komplizierten Instrumente, die zur Erhebung egozentrierter Netzwerkdaten eingesetzt werden, auch am Telefon funktionieren und diese kostengünstige Erhebungsmethode letztlich auch bessere Ergebnisse liefert.

Bei der standardisierten Befragung, insbesondere der persönlich-mündlichen Befragung kommt dem Verhalten des Interviewers eine zentrale Bedeutung zu, da es im Gegensatz zur Abfolge der Items nur bedingt kontrolliert werden kann. Die Größe dieses Verzerrungspotentials hängt auch von der Formulierung der Items ab: Bei sensitiven, nicht faktenorientierten, schwierigen oder offenen Fragen ist die Abhängigkeit vom sozialen Geschehen, das heißt der Interpretationsleistung des Interviewers für die/den Befragte/n größer. Namensgeneratoren sind in aller Regel solche schwierigen Fragen, bei denen die Interviewer/innen noch dazu über eine hohe Aufmerksamkeit und Nachfragetechniken verfügen müssen, um die gewünschte Informationsdichte zu erreichen. In verschiedenen Studien wurden Intraklassen-Korrelationen zwischen 0,15 und 0,21 errechnet, die sich auch nach der Kontrolle verschiedener Interviewer-Merkmale nur geringfügig reduzierten. Interessante Aspekte werden bei Wolf beschrieben, die sich auf die Durchführung von Studien mit oder ohne Beziehung kommerzieller Unternehmen zur Datenerhebung beziehen. Obwohl die Interviewer/innen über wesentlich mehr Erfahrung als interessierte Studierende mit sich bringen, wüssten diese auch, wie die Interviews „abgekürzt“ werden könnten, ohne dass vertragliche Verpflichtungen verletzt werden. Da es bei Netzwerkanalysen vom Nachfragen abhängt, haben Studierende hier ein dichteres Informationsniveau erreichen können, welches sich letztlich auch auf die Testgüte positiv auswirkte. Wolf zieht nun mehrere Studien heran, die beweisen, dass die Kontrolle der Interviewer/innen, praktisch umgesetzt in Interviewer-Trainings, ausschlaggebend für den Messerfolg ist.

Dabei kommt es darauf an, den Spielraum des Interviewers weitestgehend einzuschränken. Dies kann einerseits durch wiederholte Schulungen geschehen (z.B. nach Einführung der Feldphase, nach dem ersten Interview), durch laufende Kontrolle der Interviews und Nachschulungen oder durch Einschränkungen des Instruments selbst, was vorwiegend bei computergestützten Verfahren möglich wird.

Letztlich geht Wolf auf das Problem fehlender Werte ein. Diese können die Qualität der Daten insbesondere dann beeinflussen, wenn sie nicht zufällig auftreten. Generell gilt, dass Fragen im Zusammenhang mit der Erhebung sozialer Netzwerke kaum verweigert werden. Die Namensgeneratoren werden von mehr als 99,5 % der Befragten beantwortet, die Namensinterpretatoren von mehr als 98 %.

5. Fazit

Zur besseren Einschätzung der Lage habe ich auch selbständig eine Literaturrecherche in den Datenbanken WISONET und MEDLINE durchgeführt.

1. In Wisonet waren die Suchkriterien „soziales Netzwerk“ und später „soziale Inklusion“, welche ich mit dem Suchbegriff „Reliabilität“ verbunden habe. Unter „soziales Netzwerk“ erhielt ich 28 Treffer. Vier der Arbeiten erwiesen sich als unbrauchbar. Unter diesen Treffern konnte ich bei keinem einzigen auf den Volltext zugreifen, weil die vorwiegend Zeitschriftenartikel elektronisch nicht verfügbar waren. Die dabei im ersten Schritt generierten Informationen lassen sich so zusammenfassen, dass zur Erhebung sozialer Netzwerke hauptsächlich teilstrukturierte Interviews durchgeführt wurden, aber eine Angabe des Messinstruments im Abstract nicht üblich ist. Vermutlich wurde in diesen Studien keine Begleitforschung zur Reliabilität geleistet.
2. Als zweites suchte ich nach den Evaluationsstudien zu bekannten Instrumenten zur Erforschung sozialer Netzwerke wie der Netzwerkkarte oder der Ecomap. Da in Wisonet keine relevanten Treffer erzielt werden konnten, erweiterte ich die Suche in der größten Medizindatenbank (Medline) und erhielt 29 Treffer. Ecomapping ist als Erhebungsmethode auch in den Gesundheitswissenschaften, insbesondere der Pflegewissenschaft entdeckt worden. Eine relevante Arbeit beschäftigt sich mit den Vorzügen von „bildgebenden“ diagnostischen Verfahren auch für die bevölkerungsbezogene Forschung. Leider beziehen sich aber keine Arbeiten auf die statistischen Testgütekriterien.

Angesichts dessen, dass kaum Evaluationsstudien über deutschsprachig entwickelte Instrumente im Bereich der Erforschung egozentrierter sozialer Netzwerke bestehen, sieht man von einem Methodenforschungsprojekt des ZUMA einmal ab, schlage ich vor, die Validität der Messung sozialer Netzwerke in Zukunft begleit-erforschen zu lassen.

ⁱ Linn P. L., Slinde J. A.: Determination of the significance of change between pre- and posttesting periods. *Reviews of Educational Research* 1977; 47: 121-150. Zit. n. : Steiner D. L. & Norman G. R. (1995): *Health measurement scales. A practical guide to their development and use.* Second Edition. Oxford University Press. 163-180.

ⁱⁱ Guyatt G., Walter S., Norman G.: Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987; 40: 171-178.

ⁱⁱⁱ Deyo R. A., Centor R. M.: Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis* 1986; 39: 897-906.

^{iv} Kazis L. E., Anderson J. J., Meenan R. F.: Effects sizes for interpreting changes in health status. *Med Care* 1989; 27: 178-189.

^v Steiner D. L. & Norman G. R. (1995): *Health measurement scales. A practical guide to their development and use.* Second Edition. Oxford University Press. 65.