

Ulrich Frick¹, Jürgen Rehm², Uta Thien³, Thomas Spuhler⁴

¹ Institut für sozialwissenschaftliche Computeranalysen und Methoden (ICAM), Mannheim

² Schweizerische Fachstelle für Alkohol- und andere Drogenprobleme (SFA), Lausanne

³ Biometrisches Zentrum für Therapiestudien (BZT), München

⁴ Bundesamt für Statistik –Sektion Gesundheit– (BfS), Bern

Zur Konstruktion und Validierung eines Indikators für Alkoholprobleme in der schweizerischen Gesundheitsbefragung

Zusammenfassung

Ein vorgeschlagener Indikator für Alkoholprobleme aus fünf Items wurde über die üblichen klassisch-testtheoretischen Masse hinaus psychometrisch untersucht und validiert. Der Einsatz dieser Skala in Gesundheitsbefragungen der Allgemeinbevölkerung ist gerechtfertigt und bietet eine zuverlässige und sehr ökonomische Methode zur Messung von alkoholkonsuminduzierten Problemen. Für die schweizerische Gesundheitsbefragung wird daher der Einsatz dieser Skala auch für die künftigen Befragungswellen empfohlen. Der jeweilige Befragungskontext muss bei der Interpretation des Summenwertes im Einzelfall berücksichtigt werden.

Seit 1975 führt die Schweizerische Fachstelle für Alkohol- und andere Drogenprobleme (SFA) im 6-Jahres-Abstand repräsentative Umfragen zum Alkoholkonsum der schweizerischen Bevölkerung durch. Die Fragestellungen zu Konsummenge und -muster blieben dabei konstant, während Indikatoren potentiell problematischer Trinkgewohnheiten wegen ihres hohen Reaktanz- und Verfälschungsrisikos im Jahre 1987 eine völlige Neuformulierung fanden. Dazu wurden aus einem Pool möglicher Fragen fünf Items nach inhaltlicher Zusammengehörigkeit und nach sozialer Akzeptanz in Voruntersuchungen ausgewählt, die sich alle auf Problemtrinken im Sinne von Cahalan und Room¹ bezogen (Formulierungen der Fra-

gen und Antwortvorgaben s. statistische Verfahren zur Skalenebildung).

Abgefragt werden, – in Abhebung zu klassischen Alkoholismus-Tests (im deutschsprachigen Raum z.B. MALT², sowie MAST³, – keinerlei körperliche oder psychische Symptome. Vielmehr beziehen sich die Fragen auf soziale Reaktionen des persönlichen Umfeldes auf das eigene Trinkverhalten (Items 4 und 5) sowie auf die vom Probanden selbst vorgenommene Bewertung dieses Verhaltens (Items 1 bis 3). Die Messintention bestand in der Entwicklung eines Screening-Instruments⁴ für Alkoholprobleme, das nicht nur in klinischen Inanspruchnahmepopulationen, sondern in der Allgemeinbevölkerung einsetzbar sein sollte. Die vor-

liegende Arbeit beschreibt die Konstruktion eines Index zur Intensität von Alkoholproblemen auf Basis dieser fünf Items. Sie benutzt die Umfrageergebnisse von 1987⁵ als Lernstichprobe zur Skalenbildung. Für die Kreuzvalidierung der Skala werden Daten aus der „Methodenstudie Alkoholkonsum“ der Schweizerischen Gesundheitsbefragung 1992⁶ herangezogen (für eine Kurzbeschreibung s. a.⁷).

Daten und Methoden

Stichprobenbeschreibung

Die Lernstichprobe besteht aus $n = 1910$ mündlich befragten Probanden, die eine zweistufige, sprachregional disproportional modifizierte Quotenauswahl von 15–74jährigen Personen aus der schweizerischen Wohnbevölkerung im Jahre 1987 darstellen. Befragt wurden daraus alle Personen, die eine Alkoholkonsumhäufigkeit von mindestens einmal pro Jahr bejahten ($n = 1400^8$). Davon beantworten $n = 1333$ Personen alle 5 hier analysierten Items (95,2%).

Die Stichprobe zur Kreuzvalidierung besteht aus zwei Teilstichproben, die entweder persönlich

(n = 1097) oder telefonisch und anschliessend schriftlich (n = 1660/n = 1154) befragt wurden. Ziel dieser Methodenstudie war die Kontrolle möglicher Effekte der Befragungsmethode auf alle Angaben, die mit Alkoholkonsum in Verbindung stehen. Als Stichprobenverfahren gelangte in beiden Fällen ein zweistufiges random quota sampling aus der schweizerischen Wohnbevölkerung der 15–74jährigen zur Anwendung. Erhebungszeitraum war April bis Juni 1992. In der vorliegenden Arbeit werden die Daten der persönlichen Befragungen zur Kreuzvalidierung der 87er Ergebnisse benutzt; die Daten der den Telefoninterviews nachfolgenden schriftlichen Befragungen dienen zur Sensitivitätsprüfung. In persönlichen Interviews antworteten n = 1029 Personen (93,8%) vollständig, schriftlich äusserten sich n = 925 von 1025 nichtabstinenten Personen (90,2%) zu allen 5 gestellten Fragen.

Statistische Verfahren zur Skalenbildung

Frageformulierungen und Antwortvorgaben lauten wie folgt:

Item 1: „Haben Sie schon daran gedacht, weniger Alkohol zu trinken?“

Item 2: „Und haben Sie schon das Gefühl gehabt, dass Ihnen Alkohol nicht gut tut?“

Item 3: „Kommt es manchmal vor, dass Sie mehr trinken, als Sie eigentlich wollten?“

Item 4: „Haben Sie wegen Ihres Alkoholkonsums auch schon Schwierigkeiten mit anderen Leuten bekommen?“

Item 5: „Hat Ihnen schon einmal jemand gesagt, Sie sollten weniger trinken?“

Zwei der fünf Items (Item 1 und Item 4) konnten dichotom nur mit „nein“ oder „ja“ beantwortet werden. Für die anderen drei Items waren die Antwortkatego-

rien „nein“ (numerisch als 0 kodiert), „ja, aber selten“ (als 1 kodiert) und „ja, manchmal“ (mit 2 kodiert) vorgegeben.

Im ersten Analyseschritt wird daher zum einen geprüft, ob für die „diagnostische Beschreibung“ der Probanden mit Hilfe dieser fünf Items von einer qualitativen, latenten Personenvariable auszugehen ist, – mithin also die 5 „Problemitems“ lediglich eine Einteilung der Personen in qualitativ unterschiedliche diagnostische Problemklassen erlauben –, oder ob – wie erhofft – eine eindimensionale latente Personenvariable auffindbar ist, auf der die Probanden quantitativ angeordnet werden können. Zum anderen wird in diesem ersten Analyseschritt überprüft, ob die erhoffte Ordinalität der Antwortkategorien bei den Items 2, 3 und 5 empirisch abgesichert werden kann.

Die Methode der Wahl zur Beantwortung dieser beiden Fragen ist eine Latente-Klasse-Analyse (LCA, für die Anwendung in der Medizin⁹) für mehrkategorielle Items unter Berücksichtigung der Reihenfolge der Itemkategorien, wie sie von Rost¹⁰ vorgestellt wurde. Die formale Darstellung der überprüften Modelle findet sich bei¹¹. Für die weitere Darstellung der Skalenkonstruktion ist lediglich das Verständnis der Intuition dieser Verfahren notwendig:

In der LCA werden die Antworthäufigkeiten der Probanden auf den Items rekonstruiert aus den mehr oder weniger ausgeprägten Zugehörigkeiten jeder Person zu jeder aus einer möglichst kleinen Anzahl von latenten Klassen. Die idealtypischen Antwortmuster dieser latenten Klassen werden aus den empirischen Daten geschätzt, genauso wie die Anzahl zu schätzender Klassen an den Daten entschieden wird. Die Kategorien jedes Items werden als geordnet im Sinne steigender Merkmalsausprägung pro Item aufgefasst, so dass eine Darstellung der berechneten

latenten Klassen im „Merkmalsraum“ des Tests möglich wird.

Rost¹¹ hat darauf hingewiesen, dass eine überschneidungsfreie Anordnung dieser Profile der latenten Klassen eine notwendige Voraussetzung dafür ist, Aussagen über Personenmerkmale quantitativ im Sinne von „latent traits“ treffen zu können. Unter „latent traits“ werden dabei Merkmale verstanden (wie z.B. Problemtrinken), die nicht direkt messbar sind, aber indirekt aus gemessenen Indikatoren über Zuordnungsregeln erschlossen werden können. „Quantitativ“ bedeutet im vorliegenden Zusammenhang, dass die entsprechenden latenten Variablen Intervallskalenniveau aufweisen, d.h. für das Beispiel, dass hinsichtlich Problemtrinken Unterschiede zwischen Individuen quantifiziert werden können (und nicht nur eine Rangreihe aufgestellt werden kann).

Die Ordnung der Kategorien innerhalb eines Items kann überprüft werden durch eine Inspektion der geschätzten Übergangswahrscheinlichkeiten zwischen benachbarten Kategorien eines Items (bei drei Antwortvorgaben sind zwei solcher Schwellenparameter notwendig). Von Masters¹² stammt die Forderung, dass die Schwellenparameter jedes ordinalen Items entlang der Merkmalsausprägung absteigend geordnet sein müssen, wenn sie jeweils steigende Intensität des Merkmales abbilden sollen.

Für den Fall, dass es aufgrund der Ergebnisse der Latente-Klassen-Analyse des ersten Schritts sinnvoll erscheint, eine quantitative und im messtheoretischen Sinne homogene Dimension von „Alkoholproblemen“ als gemeinsame latente Variable aller 5 Items in einem zweiten Schritt zu etablieren, soll dazu eine Skalenanalyse nach Mokken¹³ durchgeführt werden.

Das Mokkenmodell geht davon aus, dass alle Items und alle Personen auf einem gemeinsamen

Eigenschaftskontinuum (hier „Alkoholprobleme“) angesiedelt werden können. Die Stellung der Items auf diesem Kontinuum wird gemessen über die relative Häufigkeit der Item-Bejahung. Die Position von Personen auf dieser Dimension ist bestimmbar über die Anzahl der bejahten Items und drückt ihre Merkmalsausprägung aus. Eine stärkere Merkmalsausprägung der Person führt zu einer höheren Chance, auch Fragen zu bejahen, die nur von relativ wenigen Personen bejaht werden.*

Es gibt keine sichere Möglichkeit, in Kenntnis der Merkmalsausprägung einer Person exakt die von ihr bejahten Items zu prognostizieren. Der Zusammenhang zwischen Merkmalsausprägung und „Lösungswahrscheinlichkeit“ ist lediglich probabilistisch, mehr oder weniger eng, je nach Test (daher auch der Name „probabilistische Testtheorie“ für diese Art von Skalenmodellen).

Das Mokkenmodell lässt die Vorstellung über die mathematische Form dieses Zusammenhangs weitgehend offen. Festgelegt ist lediglich, dass bei steigender Merkmalsausprägung die Wahrscheinlichkeit zum Bejahen auch „schwierigerer“ Fragen niemals sinken, meistens aber steigen wird („monotone, nichtsinkende Funktion“). Es besitzt die besonders günstigen Modelleigenschaften der „spezifischen Objektivität“¹⁴ und „lokalen stochastischen Unabhängigkeit“¹⁵, wenn es der Forderung der „double monotonicity“ folgt.

„Double monotonicity“ bedeutet, dass die Kurvenform, wie bei

steigender Eigenschaftsausprägung die Lösungswahrscheinlichkeit eines speziellen Items ansteigt, zwar für jedes Item des Tests offengelassen wird, dass aber gesichert ist, dass sich die Kurven aller Items (item characteristic curves) niemals überschneiden. Gilt dies, dann sind für alle Personen der Stichprobe – egal, welche Merkmalsausprägung sie auf der zu messenden Dimension besitzen –, alle Items in einer gleichen Schwierigkeitsabfolge geordnet. Dies wird unter dem Begriff der „spezifischen Objektivität“ beschrieben und bildet eine notwendige und hinreichende Voraussetzung dafür, die Personen ausschliesslich über die Summe der von ihnen bejahten Items optimal vergleichen, d.h. „messen“ zu können. Unter „lokaler stochastischer Unabhängigkeit“ wird verstanden, dass die Tatsache, ein bestimmtes Item bejaht zu haben, in keinem Falle die Lösungswahrscheinlichkeit eines anderen Items beeinflusst. Lerneffekte oder Stimmungstransfer bei der Fragenbeantwortung sind also ausgeschlossen.

Die Prüfung einer Mokkenskala erfolgt nicht über spezielle statistische Tests, sondern durch die Inspektion einer Reihe von Zusammenhangskoeffizienten zwischen den Items¹⁶, deren Definition im Anhang wiedergegeben ist. Zusätzlich muss eine Matrix der bedingten, bivariaten Antwortwahrscheinlichkeiten (sog. P-Matrix) der Skalenitems, die anhand der marginalen Lösungswahrscheinlichkeiten der einzelnen Fragen des Tests geordnet wurde, auch in ihrem Inneren monoton geordnet sein, wenn „double monotonicity“ gegeben ist¹⁷.

Zur Schätzung der LCA-Modelle wurde das Programm LACORD¹⁸, für die Berechnungen der Modellgeltungskontrollen im Rahmen der Mokkenskalierung wurde das Programm MSP¹⁹ eingesetzt.

Ergebnisse

Latente-Klassen-Analyse

Die LCA-Modelle wurden geschätzt unter der Restriktion, dass die Antwortvorgaben jedes einzelnen Items in sich geordnet sein sollen. Weitere Restriktionen wurden den Itemparametern zunächst nicht auferlegt. Verglichen wurden die Lösungen mit genau einer latenten Klasse bis hin zu sechs latenten Klassen als Personenvariable. Zusätzlich wurde überprüft, ob die Konstantsetzung der Item-Schwellenparameter jedes Items über die latenten Klassen hinweg eine günstigere Anpassung des LCA-Modelles an die Daten ermöglicht. Zum Vergleich der berechneten 12 Modelle wurde das „best information criterion“ (BIC)²⁰ herangezogen.

Das bestangepasste Modell (geringster BIC-Wert) in der Lernstichprobe wies drei latente Klassen auf und erforderte keine spezifische Schätzung der Item-Schwellenparameter in den verschiedenen Klassen: Die Übergänge zwischen benachbarten Kategorien (jeweils unter der Bedingung, dass eine der beiden Kategorien als Antwort gewählt wurde), sind pro Item für alle drei Personenklassen gleich leicht oder schwierig. Abbildung 1 stellt die drei latenten Klassen dieses Modells dar. Die Kurvenverläufe symbolisieren den Erwartungswert der Merkmalsausprägung jeder latenten Klasse über die fünf Items hinweg.

Die latente Klasse 1 umfasst 58% der Stichprobe und zeigt auf allen Items kaum einen Hinweis auf Alkoholprobleme: Die Kurve verläuft nahe beim Wert 0. Die zweite latente Klasse ist kleiner (sie umfasst rund 35% der Stichprobe) und erreicht mittelstarke Tendenz zur Zustimmung bei allen Fragen. In dieser Personengruppe ist durchaus schon von vereinzelt alkoholbezogenen Problemen zu

* Historisch stammen die entsprechenden Konzeptionen der psychologischen Testtheorie aus Wissens- und Fähigkeitstests. Aus dieser Zeit haben sich die Fachtermini „Schwierigkeit“ von Items und „Fähigkeit“ von Personen¹⁹ erhalten. Wegen falscher Konnotation werden sie im vorliegenden Text jeweils umschrieben.

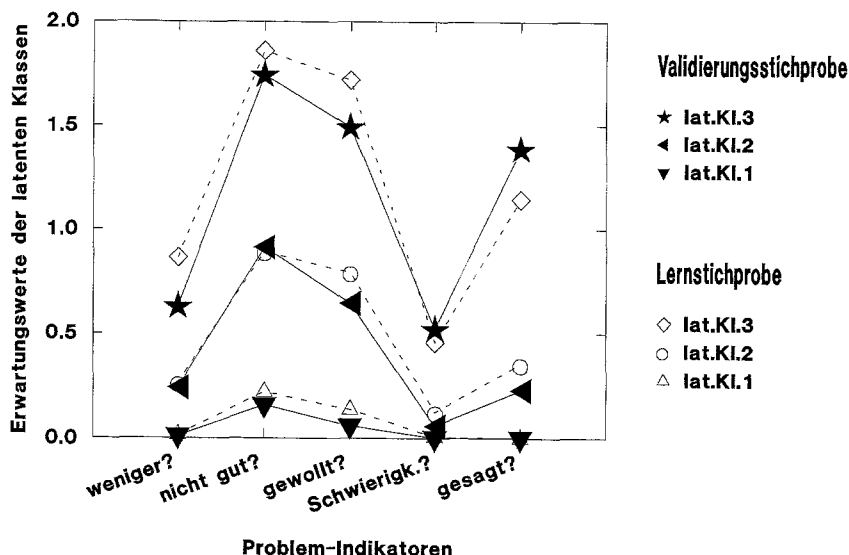


Abbildung 1. Erwartungswerte der latenten Klassen in Lern- und Validierungsstichprobe.

sprechen. Die kleinste Klasse ist die dritte: Nur 7% der Stichprobe sind hier einzuordnen. Der Erwartungswert auf allen fünf Problemfragen ist jeweils sehr hoch; augenscheinlich sind in dieser Klasse Personen mit manifesten Alkoholproblemen versammelt.

Die drei Profile der Erwartungswerte überschneiden sich an keinem Punkt. Daraus lässt sich ableiten, dass zumindest die drei latenten Klassen sich ordinal auf einem quantitativen Kontinuum von „Alkoholproblemen“ anordnen lassen. Weitergehende Analysen hinsichtlich einer kontinuierlichen Personenvariable im Sinne eines latent trait „mehr oder weniger Alkoholprobleme“ können also sinnvollerweise angegangen werden.

Zur Frage der Geordnetheit der Kategorien in den Items 2, 3 und 5 ist bei den Schätzungen der Schwellenparameter eine unterschiedliche Richtung der Antwortreihung zwischen den drei trichotomen Items zu verzeichnen. Dies bedeutet, dass die Abfolge von „nein“ zu „ja, aber selten“ und „ja, manchmal“ von den Probanden je nach Item unterschiedlich wahrgenommen wird. Ein solcher Effekt

liegt entgegen den Anforderungen an quantitativ abstufbare Antwortalternativen von Items. Wir vermuten, das mit der Formulierung „ja, aber selten“ dem Häufigkeitskonzept der Antwortkategorien durch das Wort „aber“ eine zweite Dimension sozialer Erwünschtheit hinzugefügt wurde, die einen einheitlichen Gebrauch dieses response set verhinderte. In den weiteren Analysen wurde daher die Unterscheidung von „ja, aber selten“ und „ja, manchmal“ aufgegeben und die betreffenden Items rekodiert auf die beiden Werte 0 = nein und 1 = ja.

Mokken-Analyse

Tabelle 1 informiert über die zehn Skalierbarkeitskoeffizienten H_{ij} zwischen allen paarweisen Verknüpfungen der fünf Skalennitems, über die fünf Skalierbarkeitskoeffizienten H_i , die den Zusammenhang der fünf Einzelitems jeweils mit der gesamten Skala ausdrücken, und über den globalen Koeffizienten H der getesteten Skala.

Alle Mindestkriterien für eine Mokkenskala werden erfüllt: Jeder einzelne bivariate Skalierbarkeits-

koeffizient H_{ij} unterscheidet sich signifikant und positiv von 0 (ablesbar an der Statistik Δ^* , die approximativ einer Standardnormalverteilung folgt). Die Koeffizienten H_i des Zusammenhangs der Einzelitems mit der restlichen Skala liegen alle deutlich über dem Wert von 0,4; die Mindestanforderung ist hier 0,3. Der Gesamtkoeffizient der Skalierbarkeit erreicht den Wert $H = 0,47$ und kann damit als „guter Durchschnitt“ im Hinblick auf die Skalenqualität bezeichnet werden (Mindestanforderung $H > 0,3$; ab $H > 0,5$ kann von einer „strong scale“ gesprochen werden).

Über die Skaleneigenschaft der „double monotonicity“ informiert Tabelle 2, welche die bedingten bivariaten Antwortwahrscheinlichkeiten aller Items enthält. In der Diagonale dieser sogenannten P-Matrix stehen die geschätzten (zur Methode¹³) „Stabilitätskoeffizienten“ der Items, also der erwartbare Anteil derjenigen Personen, der Item i in einer ersten Testdurchführung gelöst hat, und dies bei einer Replikation des Tests vermutlich wieder tun wird. Die Matrix dieser bedingten Lösungswahrscheinlichkeiten ist spaltenweise von oben nach unten und zeilenweise von links nach rechts aufsteigend geordnet. Lediglich innerhalb Item 4 ergibt sich beim Übergang von der bedingten Lösungswahrscheinlichkeit des Items 5 zu Item 1 eine zahlenmäßig geringe Verwerfung dieser Ordnung, sowie innerhalb Item 5 beim Übergang von Item 3 zu Item 2.

Die Größenordnung der feststellbaren Abweichungen (von der Annahme nichtsinkender Monotonie) erscheint angesichts der Stichprobengröße (die fragliche Differenz bezieht sich z. B. bei Item 4 auf einen Unterschied von $n = 4$ Personen bei einer Stichprobengröße von $n = 1333$) als vernachlässigbar²¹. Die Reliabilität der aus den fünf dichotomen Items gebildeten Skala erreicht, gemessen

Variable	Iteminhalt	% positive Antworten	H _i	Δ*
Item 1	„... schon daran gedacht, weniger Alkohol? ...“	0,16	0,46	21,86
Item 2	„... Gefühl gehabt, ... Alkohol nicht gut tut?“	0,39	0,47	21,00
Item 3	„... mehr trinken, als Sie eigentlich wollten?“	0,35	0,46	21,69
Item 4	„... Schwierigkeiten mit anderen ... bekommen?“	0,08	0,46	17,47
Item 5	„... jemand gesagt, ... sollten weniger trinken?“	0,14	0,48	22,02
Gesamtskala	mittlere Zustimmung:	0,22	H = 0,47	32,75
H _{ij}	Item 2	Item 3	Item 4	Item 5
Item 1	0,62	0,49	0,36	0,37
Item 2		0,36	0,64	0,47
Item 3			0,48	0,67
Item 4				0,42

Tabelle 1. Loevinger's H-Koeffizienten der Alkohol-Problemskala (Lernstichprobe).

Variable	Item 4	Item 5	Item 1	Item 3	Item 2
p _i	0,081	0,135	0,163	0,347	0,389
p _{ij}					
Item 4	0,031	0,041	0,038	0,053	0,063
Item 5		0,057	0,064	0,106	0,092
Item 1			0,072	0,108	0,125
Item 3				0,195	0,212
Item 2					0,230

Tabelle 2. P-Matrix der bedingten Antwortwahrscheinlichkeiten (Lernstichprobe).

über Cronbach's α , einen Wert von 0.66, bzw. gemessen über Mokken's ρ den Wert 0.68. Zusammenfassend kann also festgestellt werden, dass die fünf Fragen zu Alkoholproblemen in der Lernstichprobe psychometrisch günstige Eigenschaften aufwiesen: Befriedigende Kennwerte im Sinne der klassischen Testtheorie, Modellgeltung einer doppelt monotonen Mokkenskala seitens der probabilistischen Testtheorie.

Kreuzvalidierung

Wie bei allen statistischen Prüfungen stellt sich auch bei der

Kreuzvalidierung eines Skalierungsmodelles an einem von der Lernstichprobe unabhängigen Datenkörper das Problem, zwischen „statistischer Signifikanz“ möglicher Unterschiede bei den Modellgeltungskontrollen und inhaltlich bedeutsamen Effektgrößen ebensolcher Unterschiede zu differenzieren. Für die probabilistische Testtheorie sind dazu bislang allenfalls erste Anfänge zur Lösung vorgelegt worden²², im speziellen Fall der Etablierung einer Mokkenskala scheidet wegen der oben beschriebenen Prüfmethode ein statistischer Test von vorneherein aus.

Eine Entscheidung über die Beibehaltung des in der Lernstichprobe gewählten Skalenmodells auch für die Validierungsstichprobe erfolgte daher ausschliesslich über deskriptive Masse. Es wurde das gesamte Procedere wie unter den Rubriken Latente-Klassen-Analyse und Mokken-Analyse beschrieben an den Daten der Stichprobe der „Methodenstudie Alkoholkonsum“ (Teil: persönliche Befragung) wiederholt und sämtliche in der Lernstichprobe vollzogene Entscheidungen mit denjenigen verglichen, die in der Validierungsstichprobe zu ziehen waren. Bei der Modellauswahl in der Latenten-Klassen-Analyse wurden dieselben 12 Modelle geschätzt wie in der Lernstichprobe. Die Auswahl des bestangepassten Modelles mittels minimalen BIC-Werten führte zur Akzeptanz desselben Modelles mit drei latenten Klassen und Konstanthaltung der item-spezifischen Schwellenparameter über die latenten Klassen hinweg.*

* Eine detaillierte Dokumentation der Modellkennwerte ist auf Anfrage erhältlich beim Erstauteur.

Zur Beurteilung der Ähnlichkeit beider Analysen über die Stichproben hinweg sind in Abbildung 1 die Ergebnisse der Latente-Klassen-Analyse in der Validierungsstichprobe denjenigen der Lernstichprobe gegenübergestellt. Die Übereinstimmung ist evident. Zusätzlich nähern die relativen Grössen dieser latenten Klassen die Werte der Lernstichprobe gut an (LC1 = 67%; LC2 = 25%; LC3 = 8%).

Dieselben Probleme mit divergierenden Reihungen der Schwellenparameter bei den trichotomen Items traten wie in der Lernstichprobe auf, so dass auch in der Validierungsstichprobe eine Dichotomisierung dieser Items eine ableitbare Forderung an die Daten vor einer weiteren Skalenbildung darstellte. Die Annahme einer quantitativen Personenvariable für den Test erschien auch in der Validierungsstichprobe sinnvoll, da sich die Kurven der Erwartungswerte der drei latenten Klassen ohne Überschneidung anordnen (vgl. Abb. 1).

Eine anschliessend durchgeführte Skalenanalyse nach Mokken erbrachte bezüglich der bivariaten Skalierbarkeitskoeffizienten H_{ij}

sämtlich Werte, die grösser als 0.35 lagen (vgl. Tab. 3) und sich signifikant von 0 unterschieden. Die Item-Skalen-Koeffizienten H_i lagen für alle Items über 0.4. Der Gesamtkoeffizient H erreichte einen Wert von 0.51 („strong scale“). Die P-Matrix der bivariaten bedingten Lösungswahrscheinlichkeiten führt zu den gleichen Schlüssen wie in der Lernstichprobe: Auch in der Validierungsstichprobe kann die Annahme der „double monotonicity“ der aus den 5 Items gebildeten Skala beibehalten werden.

Die unbedingten Lösungswahrscheinlichkeiten der fünf Items nehmen annähernd dieselben Werte wie im Lernsample an und reihen die Fragen identisch nach „Schwierigkeit“ (vgl. Abbildung 2). Die Reliabilität der „Alkoholproblemskala“ in der Validierungsstichprobe beträgt nach Cronbach's α 0.68, nach Mokken's ρ 0.71.

Insgesamt kann von einer ausserordentlich hohen Stabilität der Ergebnisse der Skalierungsversuche zwischen den beiden Stichproben gesprochen werden. Die Bewährung der in der Lernstichprobe gezogenen Schlüsse zur

Skalierung kann über die Validierungsstudie als in vollem Umfang geleistet gelten.

Sensitivitätsanalyse

Sensitivitätsanalysen finden im Bereich der Biostatistik und Epidemiologie in den letzten Jahren immer mehr Verbreitung, um die Robustheit der verwendeten Methodik zu prüfen²³. Im vorliegenden Fall wurde eine Sensitivitätsanalyse hinsichtlich möglicher Veränderungen der Skalierbarkeit von Alkoholproblemen in Abhängigkeit von der Befragungsmethode (persönliche versus schriftliche Befragung) durchgeführt. Mit anderen Worten, es sollte geprüft werden, ob sich eine testtheoretisch gut bewährte, psychometrische Skala in einen anderen situationalen Kontext übertragen lässt.

In den beiden bisher dargestellten Stichproben war Grundlage der Datengewinnung ein persönliches Interview mit der Auskunftsperson. Die Fragen zu Alkoholproblemen waren dabei eingebettet entweder in eine Umfrage zum individuellen Gesundheitsverhalten⁵, oder in eine Omnibus-Mehr-

Variable	Iteminhalt				% positive Antworten	H_i	Δ^*
Item 1	„ ... schon daran gedacht, weniger Alkohol? ...“				0,14	0,44	19,02
Item 2	„ ... Gefühl gehabt, ... Alkohol nicht gut tut?“				0,35	0,54	19,86
Item 3	„ ... mehr trinken, als Sie eigentlich wollten?“				0,26	0,50	20,76
Item 4	„ ... Schwierigkeiten mit anderen ... bekommen?“				0,06	0,55	17,69
Item 5	„ ... jemand gesagt, ... sollten weniger trinken?“				0,13	0,52	21,81
Gesamtskala	mittlere Zustimmung:				0,19	$H = 0.51$	31,08
H_{ij}	Item 2	Item 3	Item 4	Item 5			
Item 1	0,57	0,44	0,35	0,39			
Item 2		0,45	0,66	0,61			
Item 3			0,68	0,55			
Item 4				0,55			

Tabelle 3. Loevinger's H-Koeffizienten der Alkohol-Problemskala (Validierungsstichprobe).

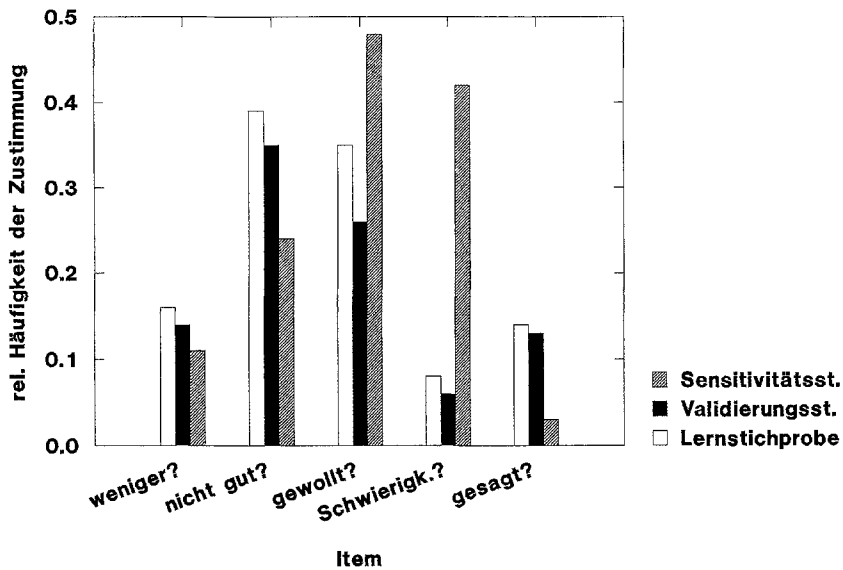


Abbildung 2. Relative Häufigkeit der Bejahung in den unterschiedlichen Stichproben.

themenumfrage des durchführenden Meinungsforschungsinstituts. In der Stichprobe zur Sensitivitätsanalyse erfolgte die Kontaktaufnahme mit den Probanden zunächst durch ein kurzes telefonisches Interview. Im unmittelbaren Anschluss daran erhielten die Befragungspersonen den gesamten Fragebogen der Schweizerischen Gesundheitsbefragung postalisch zugesandt, so dass auch hier die Beantwortung der alkoholbezogenen Items nicht isoliert, sondern in einen grösseren Fragenkomplex eingebettet stattfand.

Für diese dritte Stichprobe wurde eine erneute Skalenanalyse nach dem Mokkenmodell durchgeführt. Dabei erreichten die deskriptiven Kennwerte der „Alkoholproblemskala“ ausnahmslos zufriedenstellende Eigenschaften, die völlig vergleichbar zu den Analysen in den bisher dargestellten Stichproben lagen: Alle 10 bivariaten Skalierbarkeitskoeffizienten H_{ij} unterschieden sich signifikant von 0 und erreichten positive Werte von >0.30 . Die Item-Skala-Koeffizienten H_i erreichten sämtliche Werte von >0.4 , der Gesamtkoeffi-

zient H lag bei 0.48. Die Reliabilität der „Problemskala“ in der schriftlichen Befragungsgruppe berechnete sich über Cronbach's α zu 0.60, über Mokken's ρ zu 0.66. Das Kriterium „doppelter Monotonizität“ konnte, – wie aus der P-Matrix in Tabelle 4 ersichtlich –, auch in der Sensitivitätsstichprobe als erfüllt gelten.

Ein überraschender Effekt zeigte sich jedoch bei der Inspektion der unbedingten Lösungswahrscheinlichkeiten der fünf Items. Sie unterscheiden sich in starkem Mass von den bisherigen Ergebnissen: Während bei persönlicher Befragung nur 6% bzw. 8% bejahen, wegen ihres Alkoholkonsums schon Schwierigkeiten mit anderen Leuten bekommen zu haben, bejahen in schriftlicher Befragung dasselbe Item rund 42%. Das für die Personen der persönlichen Interviews „leichteste“ Item (Bejahung 35% bzw. 39%), die Frage nach dem Gefühl, ob Alkohol „nicht gut tut“, wird von den schriftlich befragten Personen nur mehr zu 24% bejaht und liegt bezüglich der „Leichtigkeit“ erst an dritter Stelle. In Abbildung 2

sind die Schwierigkeitsgrade der fünf Items in den drei Stichproben noch einmal vergleichend gegenübergestellt.

Insgesamt zeigt sich, dass in persönlichen Interviews weniger Probleme berichtet werden (durchschnittliche Bejahungshäufigkeit aller 5 Items 1992: 19%; 1987: 22%) als in schriftlicher Befragung (durchschnittliche Bejahungshäufigkeit 1992: 26%). Dabei unterscheidet sich die Spannweite (Differenz zwischen der seltensten und der häufigsten Bejahung eines Items) erheblich: persönlich 1992 = 29%; 1987 = 31%; schriftlich 1992 = 45%. Die schriftliche Befragung führt zu einer extremen Beantwortung an beiden Polen der Skala.

Über eine Varianzanalyse wurde der oben beschriebene Effekt der höheren Problemintensität bei schriftlicher Befragung zufallskritisch bestätigt ($F = 102.1$; $df = 1.2079$; $p \ll 0.01$). Eine Wechselwirkung zwischen Befragungsmethodik und Geschlecht des Befragten konnte ausgeschlossen werden ($F < 1$; n.s.). Die inhaltliche Bedeutung der „schriftlichen“ wie der „mündlichen“ Alkoholproblemskala bedarf daher einer intensiven Betrachtung.

Diskussion

Die fünf dargestellten Items wurden über Latente-Klassen-Analyse und Mokken-Skalierung als im messtheoretischen Sinne homogene Skala etabliert. Eine Kreuzvalidierung bestätigte die Qualität der Items als verlässliches Instrument zur Messung von Alkoholproblemen. Eine Sensitivitätsanalyse derselben Fragen unter geändertem Befragungsmodus führte erneut zu guten testtheoretischen Kennwerten. Jedoch unterschieden sich sowohl die Reihung der Items als Indikatoren für Alkoholprobleme, als auch die durchschnittliche Zustimmungshäufigkeit, als auch

Variable	Item 5	Item 1	Item 2	Item 4	Item 3
p_i	0,011	0,114	0,242	0,420	0,483
p_{ij}					
Item 5	0,011	0,018	0,022	0,024	0,023
Item 1		0,047	0,066	0,087	0,081
Item 2			0,112	0,154	0,186
Item 4				0,256	0,297
Item 3					0,330

Tabelle 4. P-Matrix der bedingten Antwortwahrscheinlichkeiten (Sensitivitätsstichprobe).

die Spannweite der positiven Antworten.

Inhaltlich bedeutet dies, dass Problemtrinker (definiert über mehr als zwei Bejahungen* auf der Problemskala) in persönlichen Gesprächen mit hoher Wahrscheinlichkeit mindestens den Items zu Trinkmengenreduzierung, Beeinträchtigungen des Wohlbefindens und „Kontrollverlust“ (Items 1–3, s. Statistische Verfahren zur Skalenbildung) zugestimmt haben, während Problemtrinker in schriftlichen Befragungen sich über die Items 2, 3 und 4 konstituieren (Beeinträchtigung des Wohlbefindens, „Kontrollverlust“** und soziale Schwierigkeiten). Besonders auffällig ist die Diskrepanz in der Beantwortung von Item 4 (soziale Schwierigkeiten), mit einer Bejahung von 42% verglichen mit 6–8% in mündlichen Befragungen. Offensichtlich werden alkoholbezogene soziale Schwierigkeiten in unterschiedlichen Befragungskontexten dramatisch unterschiedlich bewertet. Die Unterschiede in Problemniveau und Spannweite stehen in

einem inhaltlichen Zusammenhang mit den Angaben zum Alkoholkonsum in derselben Stichprobe 1992⁷. Hinsichtlich des wöchentlichen durchschnittlichen Alkoholkonsums zeigte sich, dass in persönlichen Interviews erheblich grössere Mengen angegeben wurden als in der schriftlichen Befragung. Zusätzlich antworteten tendenziell mehr Personen in der persönlichen Befragungssituation in den Extrembereichen der Verteilung²⁴. Damit zeigt sich auf Populationsebene ein komplementäres Bild zwischen der Prävalenz von Trinkmenge und der Angabe von Problemen. In persönlichen Befragungen finden sich höhere Angaben zum Konsum bei gleichzeitig niedrigeren berichteten Problemintensitäten, während die schriftliche Befragung geringere Konsumerichte bei gleichzeitig höherer Problemintensität provoziert.

Mit den vorliegenden Informationen lassen sich die psychologischen Grundlagen dieses Effektes der Befragungsmethode nicht bestimmen. Dazu wären experimentelle Forschungen notwendig (in der Tradition von Schwarz und Strack, vgl.²⁵). Als erster Hinweis auf den zugrundeliegenden Mechanismus gilt es zu beachten, dass in beiden Befragungen die Fragen zum Alkoholkonsum dem Problemindex zeitlich vorangestellt waren. Es

gälte zunächst die Hypothese zu klären, ob höhere Angaben bezüglich der Trinkmenge tendenziell zu kognitiven Dissonanzen²⁶ führen, die dann über Problemverleugnung bei der Beantwortung des Index reduziert werden. Warum aber persönliche Interviews die Angabe höherer Trinkmengen zeitigten, wäre dadurch noch nicht erklärt.

Folgerungen für die Verwertbarkeit der Skala zu Problemtrinken in Bevölkerungssurveys lassen sich jedoch auch ohne Aufklärung aller determinierenden Einflussfaktoren ziehen:

Die schriftlich erhobene Skala zu „Alkoholproblemen“ stellt in der dichotomisierten Fassung ein ökonomisches und reliables Instrument dar, das zur Verwendung in Bevölkerungssurveys hervorragend geeignet ist. „Alkoholprobleme“ werden dabei vom Inhalt her über individuelle Beeinträchtigungen (inklusive eines erlebten „Kontrollverlustes“) wie über soziale Schwierigkeiten definiert. Methodologisch bestätigt die vorliegende Arbeit eindrücklich die Notwendigkeit methodischer Begleitforschung für statistische Routineerhebungen. Nur durch solche begleitende Untersuchungen lassen sich die eingesetzten Messinstrumente korrekt interpretieren.

* Dies entspricht den oberen 20% der Eichstichprobe. Details der Normwerte sind auf Anfrage über die SFA erhältlich.

** „Kontrollverlust“ wurde apostrophiert, weil mit dem entsprechenden Item natürlich nicht Kontrollverlust im Sinne der ICD oder des DSM abgeprüft wird.

Summary**Construction and validation of an indicator for alcohol problems for the Swiss Health Survey**

A proposed scale for alcohol problems which consisted of five items, was psychometrically examined and validated. Latent class analysis was used in addition to classic psychometric measures. The use of this scale in health surveys of the general population seems to be justified, as it offers a reliable and very economic method to measure alcohol induced problems. It is therefore proposed that this scale should continue to be used in future rounds of the Swiss Health Survey. The context of the questioning has to be taken into account when interpreting the summary score, as items showed different patterns of response in personal vs interview questionnaires.

Résumé**La validation de la construction d'une échelle de mesure des problèmes liés à l'alcool pour l'Enquête Suisse sur la Santé**

Une échelle de mesure des problèmes liés à l'alcool, consistant en cinq questions, a fait l'objet d'un examen de validation psychométrique. Aux mesures psychométriques classiques, on a ajouté une analyse de structure latente. L'usage de cette échelle de mesure dans le cadre de sondages sur la santé auprès du grand public semble justifié car elle est une méthode de mesure des problèmes liés à l'alcool fiable et très économique. En conséquence, nous proposons que cette échelle de mesure continue d'être utilisée dans le cadre des rondes du Sondage Suisse sur la Santé. Cependant, lors de l'interprétation du sommaire des résultats, il faut tenir compte du contexte au sein duquel l'entrevue a eu lieu. En effet, certaines questions ont présenté des profils de réponse variables entre les questionnaires administrés dans des contextes personnel et en entrevue.

Literaturverzeichnis

- 1 Cahalan D, Room R. Problem drinking among American men. New Haven: College and University Press, 1974.
- 2 Feuerlein W, Küfner H, Ringer Ch, Anton K. Der Münchner Alkoholismustest (MALT): Test Manual. Weinheim: Beltz, 1979.
- 3 Selzer ML. Michigan Alcoholism Screening Test (MAST). Am J Psychiatry 1971; 127:89–94. (Deutsche Version bei Laboratoires SAUTER, S.A., Genf)
- 4 Hennekens CH, Buring JE. Epidemiology in Medicine. Boston: Little, Brown & Company, 1987: 331 pp.
- 5 Fahrenkrug H, Müller R. Alkohol und Gesundheit in der Schweiz. Lausanne: SFA, 1989.
- 6 Polyquest. Methodenstudie Alkoholkonsum. Band 1 & 2. Bern: Polyquest, 1992.
- 7 Rehm J, Spuhler T. Measurement error in alcohol consumption: The Swiss Health Survey. European Journal of Clinical Nutrition, in press.
- 8 Publitest. Trend. Eine repräsentative Umfrage bei Männern und Frauen in der ganzen Schweiz. Tabellenbände 1 & 2. Bern: Publitest AG, 1987.
- 9 Rindskopf D, Rindskopf W. The value of latent class analysis in medical diagnosis. Statistics in Medicine 1986; 5:21–27.
- 10 Rost J. Quantitative und qualitative probabilistische Testtheorie. Bern: Huber, 1988.
- 11 Rost J. Rating scale analysis with latent class models. Psychometrika 1988; 53:327–348.
- 12 Masters GN. Measurement models for ordered response categories. In: Langeheine R, Rost J, eds. Latent trait and latent class models. New York: Plenum, 1988:11–30.
- 13 Mokken RJ. A theory and procedure of scale analysis. The Hague: Mouton, 1971.
- 14 Fischer GH. Spezifische Objektivität: Eine wissenschaftstheo-

- retische Grundlage des Rasch-Modells. In: Kubinger K, Hrsg. *Moderne Testtheorie – Ein Abriss samt neuesten Beiträgen*. München: Psychologie Verlags Union, 1988:87–113.
- 15 *Fischer GH*. Einführung in die Theorie psychologischer Tests. Bern: Huber, 1974:211 pp.
- 16 *Loevinger J*. The technique of homogeneous tests compared with some aspects of “scale analysis” and factor analysis. *Psychological Bulletin* 1948; 45:507–530.
- 17 *Sijtsma K*. Contributions to Mokken’s nonparametric item response theory. Amsterdam: Free University Press, 1988.
- 18 *Rost J*. LACORD – Latent class analysis for ordinal variables. Kiel: Institut für die Pädagogik der Naturwissenschaften, 1990.
- 19 *Debets P, Brouwer E*. MSP – a program for Mokken scale analysis for polychotomous items. Groningen: iec ProGAMMA, 1989.
- 20 *Bozdogan H*. Model selection and Akaike’s information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 1987; 52, 3:345–370.
- 21 *DeJong A, Molenaar I*. An application of Mokkens’s model for stochastic, cumulative scaling in psychiatric research *J psychiat Res* 1987; 21, 2:137–149.
- 22 *Müller-Philipp S, Tarnai C*. Signifikanz und Relevanz von Modellabweichungen beim Rasch-Modell. In: Kubinger K, Hrg. *Moderne Testtheorie – Ein Abriss samt neuesten Beiträgen*. München: Psychologie Verlags Union, 1988: 239–259.
- 23 *Dunn G*. The design of analysis of reliability studies. *Statistical Methods in Medical Research* 1992; 1:123–157.
- 24 *Müller R, Fahrenkrug H, Rehm J*. Neue Daten zum Alkoholkonsum in der Schweizer Bevölkerung. *Drogalkohol* 1993; 17:33–37.
- 25 *Rehm J, Strack F*. Kontrolltechniken. In: Herrmann T, Tack W. eds. *Methodologische Grundlagen der Psychologie*; Göttingen: Hogrefe (Enzyklopädie der Psychologie; Themenbereich B, Serie 1, Band 1) im Druck.
- 26 *Irle M*. *Lehrbuch der Sozialpsychologie*. Göttingen: Hogrefe, 1975.

Anmerkung

Die vorliegende Arbeit entstand während eines Gastwissenschaftler-Aufenthalts des Erstautors an der Schweizer Fachstelle für Alkohol- und andere Drogenprobleme (SFA), Lausanne, im Frühjahr 1993.

Korrespondenzadresse

Ulrich Frick
estimate GmbH
Königswieserstrasse 81
D-81475 München