Aus dem Institut für Gesundheits- und Rehabilitationswissenschaften der
Ludwig-Maximilians-Universität München
Kommisarischer Leiter: Prof. Dr. med. Dennis Nowak

Grafische Modelle zur Darstellung komplexer Assoziationsstrukturen von Funktionsfähigkeitsdaten

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Ralf Andreas Strobl
aus
München

Jahr
2012

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:	Prof. Dr. Eva Grill, MPH
Mitberichterstatter:	Prof. Dr. Hans-Helge Müller
	Priv. Doz. Dr. Angela Deutschländer
Mitbetreuung durch den promovierten Mitarbeiter:	-
Dekan:	Prof. Dr. med. Dr. h.c. M. Reiser, FACR, FRCR
Tag der mündlichen Prüfung:	14.08.2012

INHALTSVERZEICHNIS

ABBILDUNGSVERZEICHNIS

TABELLENVERZEICHNIS

KAPITEL 1 EINLEITUNG

1.1 Motivation

Der medizinische Fortschritt und die verbesserten Lebensbedingungen verändern zunehmend das Aufgabenspektrum der Gesundheitsversorgung. In den meisten Regionen der Welt hat sich die kurz- und langfristige Überlebenswahrscheinlichkeit von Menschen mit schwerwiegenden oder chronischen Erkrankungen oder nach einem Unfall erhöht. Beispielsweise stieg die 28-Tage Überlebensrate nach koronarer Herzkrankheit oder nichttödlichem Myokardinfarkt innerhalb von zehn Jahren von ungefähr 40% im Jahr 1993 auf 60% im Jahr 2003 [1]. Durch die erhöhte Lebenserwartung und die signifikant gestiegenen Chancen, ernsthafte Krankheiten zu überleben geht für die Betroffenen das Risiko einher, dauerhafte Einschränkungen der Funktionsfähigkeit zu erleiden und Behinderungen zu entwickeln.

Aufgabe der Gesundheitsversorgung ist daher nicht nur eine angemessene krankheitsbezogene Versorgung sicherzustellen, sondern in allen Phasen der Gesundheitsversorgung durch rehabilitative Strategien Funktionsfähigkeit zu adressieren um permanenter Behinderung vorzubeugen [2]. Der im Juni 2011 veröffentlichte Bericht 'World report on disability' der Weltgesundheitsorganisation (WHO) unterstreicht diesen Sachverhalt eindringlich [3]:

- Weltweit leben über 1 Milliarde Menschen mit Behinderung
- Zwischen 100 und 190 Millionen Menschen leben mit starker Funktionseinschränkung
- Behinderung betrifft insbesondere Frauen und Menschen mit niedrigem Lebensstandard
- Ab einem Alter von 45 Jahren ist bereits mehr als jeder 10. von Behinderung betroffen (siehe Abbildung 1.1).

Um einen Patienten umfassend zu beurteilen und adäquat zu behandeln, müssen körperliche Symptome richtig eingeschätzt, aber auch individuelle und umweltbezogene Ressourcen bestmöglich genutzt werden. Nur so ist es möglich, eine optimale Behandlung anzubieten. Die Weltgesundheitsorganisation (WHO) definiert Gesundheit als einen 'Zustand des vollständigen körperlichen, geistigen und sozialen Wohlergehens und nicht nur das Fehlen von Krankheit oder Gebrechen' [4, 5].

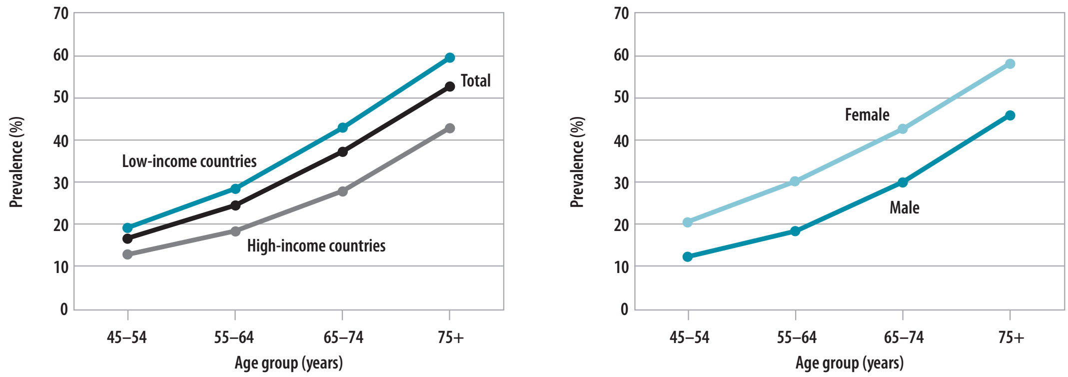

Abbildung 1.1 – *Altersspezifische Prävalenz von Behinderung für verschiedene Einkommensstufen und Geschlecht (Quelle: 'World report on disability' [3])*

Trotz der Relevanz von Funktionsfähigkeit für die Versorgung ist der Schwerpunkt der gesundheitsbezogenen Forschung nach wie vor auf die Erforschung von Mortalität und Morbidität gelegt [6]. Es fehlt ein einheitliches theoretisches Modell, um Forschung mit dem Fokus der Funktionsfähigkeit international und multidisziplinär vergleichbar durchzuführen. Mit der Verabschiedung der Internationalen Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF) im Jahr 2001 hat die WHO eine Grundlage dafür geschaffen und damit den Weg für weitere Analysen geebnet [7].

Durch die Einführung der ICF, als gemeinsame Sprache zur Beschreibung von Funktionsfähigkeit, ist es nun möglich innovative und komplexe Modelle zur Beschreibung von Funktionsfähigkeit zu entwickeln. Klinische Forschung basiert meistens auf der Untersuchung von Aktivitäten des täglichen Lebens oder der Lebensqualität [8], die unter klinischen Indizes subsumiert werden [9, 10]. Solche Indizes, meistens Summenscores, maskieren jedoch die Effekte individueller Variablen und sind meist nicht geeignet, mehrdimensionale Sachverhalte auf eine eindimensionale Größe zu reduzieren. Dies bedingt, dass viele Indizes keine validen Ergebnisse liefern können, die es erlauben, die Funktionsfähigkeit des Menschen, deren Entwicklung oder sogar die Effekte einer Intervention zu beurteilen.

Modelle, die diesem mehrdimensionalem Charakter gerecht werden, sind idealerweise in der Lage die komplexen Zusammenhänge von einzelnen Aspekten der Funktionsfähigkeit umfassender zu beschreiben als es mittels Inzidenzen und Prävalenzen realisierbar ist. Ein neuer Ansatz zur Modellierung sollte also nicht nur jeden Einzelaspekt für sich betrachten, sondern insbesondere auch die komplizierten Wirkmechanismen und Wechselspiele zwischen den einzelnen Aspekten berücksichtigen.

Anstatt ein eigenes Modell für diese Situation von Grund auf neu zu entwickeln, bietet sich eine interessante und viel versprechende Methode aus der Genforschung an. In der Genforschung werden so genannte graphische Modelle verwendet um die Zusammenhänge zwischen Genen darzustellen und damit deren Auswirkungen auf den Gesundheitszustand nachvollziehbar zu machen. Die ansprechenden Ergebnisse von graphischen Modellen in der Genforschung haben bereits dazu geführt, dass diese mehr und mehr in anderen Bereichen eingesetzt werden, z.B. in der klinischen Forschung [11]. Kalisch et. al beschrieben 2010 potentielle Anwendungsgebiete graphischer Modelle. Neben der Visualisierung von Assoziationsstrukturen dienen graphische Modelle der Identifizierung von Variablen die ein gemeinsames Konzept beschreiben. Zum Beispiel beschreiben die Variablen 'Gehen', 'Auto fahren' und 'Treppen steigen' die übergeordnete Dimension 'Mobilität'. Basierend auf solchen Variablengruppen könnte dann eine Score zur Beurteilung von Mobilität entwickelt werden [12].

Ziel dieser Dissertation ist es zu zeigen, dass sich Methoden zur Entwicklung von graphischen Modellen aus der Genforschung adaptieren lassen, um Zusammenhänge von Daten der Funktionsfähigkeit - gemessen mit Kategorien der ICF - anschaulich darzustellen.

In den folgenden beiden Abschnitten werden die beiden zentralen Konzepte dieser Arbeit genauer vorgestellt. Zum einen das Modell und das Klassifikationssystem der ICF und zum anderen graphische Modelle und die wichtigsten Bausteine die nötig sind ein graphisches Modell zu 'lesen'.

1.2 ICF - Das Modell und das System

Grundlage für das der ICF zugrundeliegende biopsychosoziale Modell (siehe Abbildung 1.2) ist die Integration des medizinischen und des sozialen Modells von Behinderung. Aus Sicht des medizinischen Modells wird Funktionsfähigkeit und Behinderung als eine Eigenschaft der erkrankten Person verstanden, welches in Folge eines akuten Ereignisses oder einer Krankheit auftritt. Das soziale Modell betrachtet Behinderung als ein Problem der Gesellschaft, geschaffen durch unzureichende Integration Betroffener.

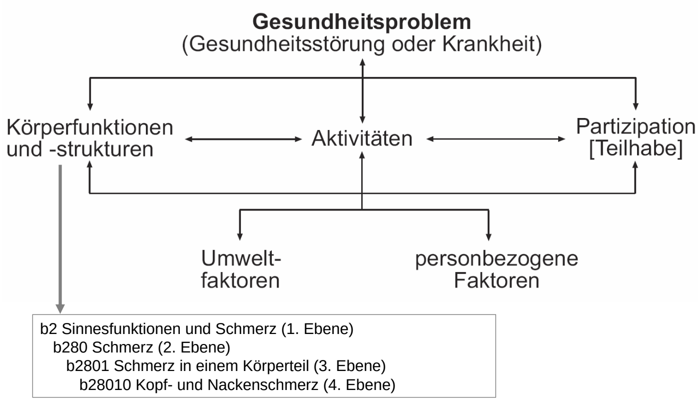

Abbildung 1.2 – *Das biopsychosoziale Modell der ICF und ein Beispiel für die Hierarchiestruktur innerhalb der ICF [7])*

Das Modell der ICF beschreibt die Funktionsfähigkeit eines Menschen als ein Zusammenspiel der Komponenten Körperfunktionen und -strukturen, Aktivitäten und Teilhabe vor dem Hintergrund eines konkreten Gesundheitsproblem (meist durch ICD kodiert) und gesundheitsrelevanten Kontextfaktoren (Umweltfaktoren und personbezogene Faktoren) (für die Definition der in der ICF verwendeten Begriffe siehe Tabelle 1.1).

Tabelle 1.1 – *Begriffsdefinitionen aus der deutschen Fassung der ICF [13]*

Begriff	Definition
Körperfunktionen	physiologische Funktionen von Körpersystemen
Körperstrukturen	anatomische Teile des Körpers
Aktivität	Durchführung einer Aufgabe oder Handlung (individuelle Perspektive)
Partizipation	Einbezogensein in eine Lebenssituation (gesellschaftliche Perspektive)
Umweltfaktoren	materielle, soziale und einstellungsbezogene Umwelt
personbezogene Faktoren	spezieller Hintergrund des Lebens und der Lebensführung

Dem Modell der ICF ist eine hierarchisch organisierte Taxonomie zugeordnet. Dieser Taxonomie zufolge detaillieren sich die Komponenten in der ersten Gliederungsebene in verschiedene Domänen, z.B. beschreiben zwei Domänen der Komponente Körperfunktionen mentale Funktionen 'b1' und Sinnesfunktionen und Schmerz 'b2'. Diese Domänen beinhalten auf der 2. Ebene Kategorien, welche die Aspekte der Funktionsfähigkeit genauer beschreiben, z.B. 'b210 Funktionen des Sehen' oder 'b280 Schmerz'. Weitere hierarchische Detaillierungen bis hin zur 4. Ebene sind möglich (siehe Abbildung 1.2). Mit Hilfe dieser Kategorien ist es denkbar, allen Aspekten der Funktionsfähigkeit, wie sie in der ICF vorgesehen sind, eine Ausprägung zuzuteilen und damit die Funktionsfähigkeit eines Menschen zu charakterisieren und gesundheitsrelevante Kontextfaktoren zu beschreiben.

1.3 Graphische Modelle

1.3.1 Aufbau von graphischen Modellen

Graphische Modelle dienen der Visualisierung von bedingten Abhängigkeiten zwischen Zufallsvariablen, d.h. statistischen Zusammenhängen zwischen Variablen unter der Bedingung, dass Informationen zum Zustand von einer oder mehreren anderen Variablen vorhanden sind (siehe Definition 1). Die bedingte Unabhängigkeit von Zufallsvariablen ist das zentrale Konzept graphischer Modelle und dessen Verständnis unabdingbar um graphische Modelle richtig interpretieren zu können.

Definition 1 (Bedingte Unabhängigkeit)
Seien X, Y, Z Realisationen von Zufallsvariablen mit gemeinsamer Verteilung. X heißt bedingt unabhängig von Y gegeben Z, kurz $X \perp Y | Z$, genau dann wenn gilt:

$$P(X|Y, Z) = P(X|Z)$$

.

Mit anderen Worten ausgedrückt: es genügt zur Vorhersage von X alle Informationen von Z zu berücksichtigen. Die Variable Y beinhaltet keinerlei zusätzlicher Informationen. Mit einem graphischen Modell ist es möglich solche Abhängigkeitsbeziehungen übersichtlich durch einen Graphen zu visualisieren [14].

Zufallsvariablen werden in graphischen Modellen durch Knoten repräsentiert. Bedingte Abhängigkeiten zwischen Zufallsvariablen werden durch Verbindungen (meist als Kanten bezeichnet) zwischen solchen Knoten dargestellt. Analog dazu repräsentiert die Abwesenheit einer Kante eine bedingte Unabhängigkeit. Man unterscheidet zwei Arten von Kanten: gerichtete und ungerichtete Kanten.

- *Gerichtete* Kanten (Pfeile) stehen für einen kausalen Zusammenhang zwischen zwei Variablen, bei der der Pfeil von der Ursache zu Wirkung gerichtet ist.
- *Ungerichtete* Kanten stehen für einen ungerichteten Zusammenhang, d.h. Ursache und Wirkung können nicht unterschieden werden.

Mit diesen Bausteinen lassen sich zwei Graphentypen erstellen. Ungerichtete Graphen, also Graphen ohne gerichtete Kanten, und sogenannte azyklisch gerichtete Graphen (Directed Acyclic Graphs - DAG), also Graphen mit gerichteten Kanten die keine Zirkelschlüsse zulassen. In Abbildung 1.3 sind beispielhaft zwei verschiedene Definitionen von Confounding, eine Scheinkorrelation zwischen Exposition und Endpunkt hervorgerufen durch eine dritte Störgröße, als graphisches Modell dargestellt: links mit ausschließlich ungerichteten Kanten und rechts als DAG mit ausschließlich gerichteten Kanten.

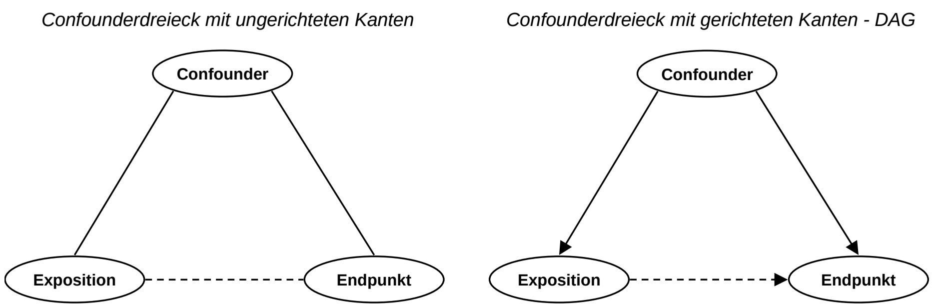

Abbildung 1.3 – *Confounderdreieck dargestellt mit gerichteten (links) und ungerichteten (rechts) Kanten)*

Grundlage zur Erzeugung gerichteter Graphen ist die Kenntnis der kausalen Zusammenhänge. Eindeutig können solche Zusammenhänge nur durch experimentelle kontrollierte Studien (z.B. RCTs) gezeigt werden. Weil diese Arbeit auf querschnittlichen Beobachtungsstudien basiert, werden ausschließlich ungerichtete Graphen untersucht.

1.3.2 Beispiel: Graphisches Modell zur Framingham-Herzstudie

Das Konzept von ungerichteten Graphen lässt sich gut am Beispiel der Framingham-Studie demonstrieren. Die Daten basieren auf einer Kohortenstudie zur Erforschung von Risikofaktoren und Umwelteinflüssen auf die Prävalenz von Herzerkrankungen und wurden bereits im Jahr 1995 von Klein et. al als Beispiel für graphische Modelle verwendet [15]. Abbildung 1.4 postuliert Abhängigkeitsbeziehungen zwischen Blutdruck und potentiellen Risikofaktoren.

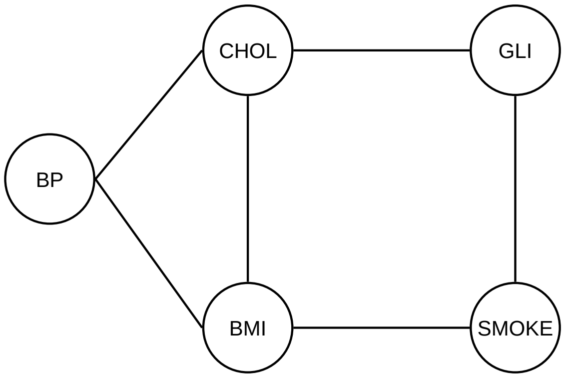

Abbildung 1.4 – *Ein mögliches graphisches Modell für den Zusammenhang von Blutdruck (BP), Cholesterin (CHOL), Body Mass Index (BMI), Glukoseintoleranz (GLI) und Rauchen (SMOKE).*

Folgende bedingte Unabhängigkeiten sind in Abbildung 1.4 dargestellt:

1. $BP \perp GLI | CHOL, BMI$
2. $BP \perp GLI | CHOL, SMOKE$
3. $BP \perp SMOKE | CHOL, BMI$
4. $BP \perp SMOKE | CHOL, GLI$
5. $BMI \perp GLI | CHOL, SMOKE$
6. $CHOL \perp SMOKE | BMI, GLI$

Aus diesem Modell könnte folgende Schlussfolgerung gezogen werden: Um Blutdruck zu prognostizieren genügt es Cholesterin und BMI zu kennen. Glukoseintoleranz und Rauchen sind mit Blutdruck lediglich indirekt assoziiert, d.h. über andere Faktoren.

1.4 Vorstellung der Beiträge

Um sich ein Gesamtbild von einem Patienten mit all seinen erlebten Problemen und Funktionsdefiziten zu machen, fehlt es an geeigneten Methoden um die komplexen Zusammenhänge anschaulich darzustellen. Die Erforschung von Interaktionen zwischen einzelnen Aspekten der Funktionsfähigkeit mithilfe statistischer Methoden stellt aufgrund der hohen Komplexität menschlicher Körperfunktionen und deren Zusammenspiel mit der Teilhabe am täglichen Leben eine große Herausforderung dar.

Um diese Herausforderung bewältigen zu können wurde eine Methode zur Entwicklung von graphischen Modellen aus der Genforschung adaptiert und deren Eigenschaften bei der Darstellung von Zusammenhängen von Daten der Funktionsfähigkeit in verschiedenen Szenarien getestet und mit anderen statistischen Verfahren verglichen.

Im Folgenden wird jeweils der Beitrag des Doktoranden zu den verfassten Fachartikeln dargelegt und die Inhalte der einzelnen Veröffentlichungen kurz vorgestellt.

1.4.1 Graphical models illustrated complex associations between variables describing human functioning

Erster Schritt der Untersuchungen war es, graphische Modelle an einem real existierenden Datensatz zu testen. Dazu wurde eine Methode von Meinshausen und Bühlmann [16] zur Identifikation von graphischen Modellen an ICF Daten angepasst. Hierbei dienen penalisierte lineare Regressionsmodellen (Least Absolute Shrinkage and Selection Operator - LASSO) [17] der Identifikation der einzelnen Kanten des graphischen Modells. LASSO minimiert die Fehlerquadratsumme einer linearen Regression unter der Nebenbedingung, dass die Summe der Regressionkoeffizienten nicht größer als ein apriorisch definierter Parameter sein darf. Je kleiner dieser Parameter gewählt wird, desto kleiner müssen diese Summe und damit die einzelnen Koeffizienten werden. Bei diesem Prozess werden unbedeutende Koeffizienten Null gesetzt und werden damit faktisch aus der Gleichung ausgeschlossen. Diese gelten damit als bedingt unabhängig von der Zielgröße. Bei der Identifikation des graphischen Modells fungiert jede Variable sowohl als Zielgröße für die anderen Variablen, als auch als Prädiktor.

Um die große Variabilität in der Modellschätzung zu reduzieren, wurde darüber hinaus ein Resamplingverfahren adaptiert - sogenanntes Bagging [18]. Beim Bagging werden aus der Originalstichprobe durch zufälliges Ziehen mit Zurücklegen neue Stichproben generiert, jeweils die LASSO-Regression berechnet und die Ergebnisse am Ende gemittelt. Die Idee, eine Synthese aus LASSO und Bagging zu verwenden, um optimale Ergebnisse zu erhalten, stammt von Francis Bach. Dieser gab der Methode auch einen eigenen Namen: Bolasso [19].

Mit dieser modifizierten Methode konnten Daten von Patienten aus der Frührehabilitation analysiert, die Ergebnisse visualisiert und diskutiert werden. Beispielhaft diskutiert wurde eine Struktur die Aspekte von Kommunikation beschreibt. Als ein zentraler Punkt der Strutkur wurde *Sprechen* (d330) identifiziert. Die Fähigkeit 'Sprechen' wird von drei separaten Aspekten beeinflusst:

1. der körperlichen Voraussetzung zum Sprechen (*Artikulationsfunktionen* (b320), *Alternative stimmliche Äußerungen* (b340) und *Funktionen der Stimme* (b310))
2. den Fähigkeiten Konversation zu betreiben (*Konversation* (d350), *Kommunikationsgeräte und -techniken benutzen* (d360), *Lesen* (d166) und *Schreiben* (d170))
3. den geistigen Voraussetzungen (angefangen bei *Kognitiv-sprachliche Funktionen* (b167) bis hin zu *Funktionen des Denkens* (b160)).

Der Doktorand war Erstautor dieser Studie und damit hauptverantwortlich für deren Konzeption, die Analyse und der Formulierung des Manuskripts.

1.4.2 Graphical modeling of binary data using the LASSO: a simulation study

In diesem Kapitel werden die einzelnen, noch unbekannten Parameter und deren Einfluss auf die Güte des graphischen Modells mittels einer Simulationsstudie beurteilt. Bei dieser Studie wurden Daten aus a-priori festgelegter Verteilung und Assoziationsstruktur erzeugt und jeweils ein graphisches Modell berechnet. Das datenbasierte graphische Modell kann dann mit dem bekanntem Modell verglichen und geeignete Gütemaße berechnet werden. Die Beurteilung der Güte der verschiedenen Parametereinstellung beruhte im Wesentlichen auf der Structural Hamming Distance (SHD) und dem Youden Index. Die SHD vergleicht die Anzahl an Schritte die nötig sind, um das datenbasierte Modell in das optimale Modell zu transformieren. Der Youden Index stellt eine Synthese aus falsch identifizierten Kanten und falsch nicht identifizierten Kanten dar und fügt der SHD damit eine weitere Dimension hinzu. Mit diesen Maßen wurde der Einfluss der verschiedenen Regularisierungsparameter (z.B. die Höhe des Bestrafungsterms der penalisierten Regression) beurteilt. In der Arbeit wurde das in Kapitel 2 beschriebene Verfahren mit alternativen Methoden zur Identifikation eines Bestrafungsterms verglichen, dem Akaike Information Criterion (AIC), dem Bayesian Information Criterion (BIC) und mittels Kreuzvalidierung. Bei einer hinreichend großen Stichprobe (n $>$ 500) zeigte sich die Bolasso-basierte Methode in allen Kriterien überlegen.

Der Doktorand war Erstautor dieser Studie und damit hauptverantwortlich für deren Konzeption, die Analyse und der Formulierung des Manuskripts.

1.4.3 Dimension reduction in human functioning and disability outcomes research: graphical models versus principal components analysis

Neben der entwickelten Methode kann auch die Hauptkomponentenanalyse (Principal Component Analysis - PCA) zur Identifizierung von Assoziationstrukturen verwendet werden. PCA fasst einzelne Variablen, die stark auf eine latente Variable laden, zu Gruppen zusammen, die als eigene Dimension verstanden werden. Im Gegensatz dazu werden bei der Bolasso-basierten Methode Gruppen von eng vernetzten Variablen als eigene Dimension behandelt. Die Ergebnisse der PCA können als graphisches Modell visualisiert und damit mit den Ergebnissen der Bolasso-basierten Methode verglichen werden.

Basierend auf einem Datensatz mit 1048 Patienten aus 14 Ländern mit Rückenmarkverletzungen wurden beispielhaft beide Methoden miteinander verglichen. Im Mittel zeigte sich eine Übereinstimmung der Variablen innerhalb der einzelnen Dimensionen von 75% . Darüber hinaus zeigte sich, dass die meisten der gefundenen Dimensionen der Taxonomie der ICF entsprechen.

Der Doktorand führte die Datenanalyse durch und war beim konzeptionellen Aufbau beteiligt. Er wies auf die Gemeinsamkeiten von Hauptkomponentenanalyse und graphischen Modellen hin und war an der Formulierung des Manuskripts beteiligt, insbesondere im Methoden- und Diskussionskapitel.

1.4.4 Graphical modeling to illustrate associations between variables describing functioning in head and neck cancer patients

Schließlich wurde die entwickelte Bolasso-basierte Methode und die gewonnenen Erkenntnisse dazu verwendet die Zusammenhänge von Aspekten der Funktionsfähigkeit gemessen mit der ICF für Patienten mit Kopf-Hals Tumoren darzustellen.

Ein Experte identifizierte zunächst relevante ICF Kategorien aus denen in der Folge ein graphisches Modell berechnet wurde. Es fand sich erneut eine aussagekräftige Struktur um die Kategorie *Sprechen* (d330). Erweitert wurde diese Struktur um typische krankheitsspezifische Aspekte aus der Komponente der Körperfunktionen und -strukturen, nämlich um *Funktionen der Nahrungsaufnahme*(b510), *Struktur des Mundes* (s320) und *Funktionen der Stimme* (b310).

Der Doktorand führte die Datenanalyse durch, korrigierte Abschnitte der Arbeit und war an der Formulierung des gesamten Manuskripts beteiligt, insbesondere im Methoden- und Ergebniskapitel.

1.5 Zusammenfassung

Bislang fehlen geeignete Instrumente, um aus den komplexen Zusammenhängen der einzelnen Komponenten der Funktionsfähigkeit eines Menschen, wie sie unter anderem mit Hilfe der Internationalen Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit (ICF) beschrieben werden können, ein angemessenes Gesamtbild zu zeichnen. Das Ziel dieser Arbeit war die Entwicklung und praktische Anwendung eines solchen Instruments basierend auf graphischen Modellen, einer statistischen Methode zur Visualisierung von komplexen Zusammenhangsstrukturen.

Um dieses Ziel zu erreichen wurden vier Studien durchgeführt, die die Anwendbarkeit von graphischen Modellen systematisch beleuchten und auf Probleme und Restriktionen dieser hinweisen.

Zunächst wurden Algorithmen zur Entwicklung von graphischen Modellen, welche ursprünglich für die Darstellung hochdimensionaler Daten in der Genforschung entwickelt wurden, adaptiert, um Aspekte menschlicher Funktionsfunktionsfähigkeit darzustellen. Die erste Publikation 'Graphical models illustrated complex associations between variables describing human functioning' legt den Hauptaugenmerk auf die Untersuchung der Machbarkeit und der Augenscheinvalidität der Ergebnisse dieser adaptierten Methode [20]. In dem gefundenen graphischen Modell zeigten sich Strukturen, die unterschiedliche krankheitstypische Aspekte widerspiegeln. Da die Augenscheinvalidität in dieser Studie ausschließlich ex-post anhand von Expertenwissen und wissenschaftlicher Literatur geprüft wurde, wäre eine Simulationsstudie darüber hinaus in der Lage die Validität der Ergebnisse objektiv zu überprüfen.

Nachdem die Machbarkeit und Anwendbarkeit der Methode auf ICF-Daten gezeigt werden konnte, untersuchte die zweite Publikation 'Graphical modeling of binary data using the LASSO: a simulation study' in einer umfangreichen Simulationsstudie die Bedeutung der verschiedenen Modelparameter für das Gesamtergebnis. Die verwendete Methode wurde mit alternativen Methoden verglichen und theoretisch fundiert [21]. Besonders in Settings mit hoher Fallzahl zeigte sich die Methode den untersuchten Alternativen überlegen und identifizierte einen Großteil der bekannten Struktur richtig.

Die dritte Publikation 'Dimension reduction in human functioning and disability outcomes research: graphical models versus principal components analysis' verglich Ergebnisse der entwickelten Methode mit denen einer alternativen Methode, der Hauptkomponentenanalyse [22]. Es zeigte sich eine hohe Übereinstimmung zwischen den Ergebnissen beider Methoden, was die Validität der Ergebnisse graphischer Modelle unterstreicht.

In der vierten Publikation 'Graphical modeling to illustrate associations between variables describing functioning in head and neck cancer patients' wurden die gewonnenen Erkenntnisse einem Praxistest unterzogen und an einem klinischen Datensatz von Patienten mit Kopf-Hals-Tumoren angewandt [23]. Es fand sich erneut eine aussagekräftige Struktur um die Kategorie *Sprechen* (d330), die um typische krankheitsspezifische Aspekte dieser Erkrankung erweitert wurden: die *Funktionen der Nahrungsaufnahme* (b510), die *Struktur des Mundes* (s320) und *Funktionen der*

Stimme (b310). Neu war hier, dass die potenziell relevanten Variablen im Vorfeld von Experten ausgewählt wurden.

Die Ergebnisse der Arbeit zeigen, dass graphische Modelle in der Lage sind, komplexe Assoziationen zwischen Aspekten der Funktionsfähigkeit zu visualisieren. Mit Hilfe solcher Modelle können, basierend auf dem Zusammenspiel statistischer Methoden und Expertenwissen, zentrale Punkte in den Assoziationsstrukturen identifiziert werden, die beispielsweise besondere therapeutische Bedeutung haben oder als Rehabilitationsziele in Frage kommen. Im Rahmen der Analyse epidemiologischer und klinischer Studien könnten diese Modelle benutzt werden, um potentielle Störgrößen (Confounder) zu identifizieren, welche bei der Analyse besonderer Aufmerksamkeit bedürfen.

1.6 Summary

Up to the present, there is a lack of methods which give an appropriate comprehensive picture of the complex association structures between single aspects of human functioning. The objective of this thesis is the development of an instrument based on so-called graphical models and the furthering of its practical applications. Graphical models are a statistical tool suited to visualize the association structures described above. To achieve this goal, four studies were conducted that systematically examine the applicability of graphic models and point out potential problems and restrictions.

The first study examined whether algorithms which were developed to analyse high-dimensional data in genetic research can be adapted to describe relationships on the macroscopic level, i.e. to describe components of human functioning. The publication 'Graphical models illustrated complex associations between variable describing humanly functioning' pays special attention to the feasibility of the method and to assessing the face-validity of the results [20]. The resulting model revealed meaningful association structures around different, but common aspects of human functioning. However, the meaningfulness and validity of the results have been deduced ex-post from expert knowledge and literature. These findings should also be confirmed in a simulation study.

Since the feasibility and applicability of the method could be shown using ICF data, the second publication 'Graphical modeling of binary data using the LASSO: a simulation study' focuses on examining the validity of the results from a technical point of view. Therefore, the question of whether the method is able to correctly identify complex structures was investigated. Different parameters for computing graphic models and their influence on the final model were examined and the proposed method was given a theoretical foundation [21]. Particularly in settings with large sample sizes, the proposed method turned out to be superior to the alternatives and it accurately identified the graphical model.

In the third publication 'Dimension reduction in human functioning and disability outcomes research: graphical models versus principal components analysis', the results of the proposed method were compared with the results of an alternative method: principal component analysis [22]. High agreement between the results of both methods could be shown. This finding further supports the validity of the results.

All these findings have been applied and practically tested in the fourth publication 'Graphical modeling to illustrate associations between variables describing functioning in head and neck cancer patients'. Here, the method was applied to clinical data describing patients with head and neck cancer [23]. The study revealed a meaningful structure around the category *speaking* (d330) and related categories that are typically affected in patients with this condition. The categories included *ingestion functions* (b510), *structure of the mouth* (s320) and *voice functions* (b310). In this study, the potentially relevant variables were selected by experts prior to the analysis.

This thesis can show that graphical models are suitable for visualizing complex associations, aspects of human functioning in particular, which are usually strongly correlated. A combination of graphical models and expert knowledge may be the basis for the identification of suitable targets, e.g. for defining rehabilitation goals or for identifying interventions of special therapeutic significance. Additionally, they might be a basis for the identification of confounders and intermediate variables, as well as the selection of parsimonious sets of variables for multivariate epidemiological modeling.

Literaturverzeichnis

[1] Hannelore Löwel, Christa Meisinger, Margit Heier, Allmut Hörmann, and Wolfgang von Scheidt. Herzinfarkt und koronare Sterblichkeit in Süddeutschland. *Deutsches Ärzteblatt*, 10(103):616–622, 2006.

[2] Ellen Nolte and Martin McKee. *Caring for People with Chronic Conditions: A Health System Perspective*. European Observatory on Health Systems and Policies Series. Open University Press, Glasgow, 2008.

[3] WHO. *World Report on Disability*. The World Bank, Malta, 2011.

[4] WHO. Constitution of the World Health Organization. *New York*, 2005.

[5] WHO. Verfassung der Weltgesundheitsorganisation. http://www.independentwho.info/Documents/ONU/constitution_OMS_DE.pdf, 2009.

[6] Walter R. Frontera, Marcus J. Fuhrer, Alan M. Jette, Leighton Chan, Rory A. Cooper, Pamela W. Duncan, John D. Kemp, Kenneth J. Ottenbacher, P Hunter Peckham, Elliot J. Roth, and Denise G. Tate. Building research capacity. Proceedings of the 2005 Rehabilitation Medicine Summit. April 28-29, 2005. Washington, D.C., USA. *Am J Phys Med Rehabil*, 84(12):913–1052, 2005.

[7] WHO. International Classification of Functioning, Disability and Health: ICF, 2001.

[8] R. Hebert, C. Brayne, and D. Spiegelhalter. Incidence of functional decline and improvement in a community-dwelling, very elderly population. *Am J Epidemiol*, 145(10):935–44, 1997.

[9] P. Raina, S. Dukeshire, J. Lindsay, and L. W. Chambers. Chronic conditions and disabilities among seniors: an analysis of population-based health and activity limitation surveys. *Ann Epidemiol*, 8(6):402–9, 1998.

[10] M. D. Patel, K. Tilling, E. Lawrence, A. G. Rudd, C. D. Wolfe, and C. McKevitt. Relationships between long-term stroke disability, handicap and health-related quality of life. *Age Ageing*, 35(3):273–9, 2006.

[11] CL Tsai and CA Camargo. Methodological considerations, such as directed acyclic graphs, for studying äcute on chronic̈disease epidemiology: Chronic obstructive pulmonary disease example. *Clinical Epidemiology*, 62:982–990, 2009.

[12] M. Kalisch, B. A. Fellinghauer, E. Grill, M. H. Maathuis, U. Mansmann, P. Buhlmann, and G. Stucki. Understanding human functioning using graphical models. *BMC Med Res Methodol*, 10:14, 2010.

[13] WHO. Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit, 2005.

[14] Ronja Foraita. *Der Conditional Synergy Index: Ein Maß zur Bestimmung einer Gen-Gen-Interaktion basierend auf graphischen Modellen*. PhD thesis, 2008.

[15] John P Klein, Niels Keiding, and Sven Kreiner. Graphical Models for Panel Studies, illustrated on data from the Framingham Heart Study. *Statistics in Medicine*, 14:1265–1290, 1995.

[16] Nicolai Meinshausen and Peter Buehlmann. High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*, 34:1436–1462, 2006.

[17] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58:267–288, 1996.

[18] L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.

[19] Francis R. Bach. Bolasso: Model consistent lasso estimation through the bootstrap. *preprint*, 2008.

[20] R Strobl, G Stucki, E Grill, M Müller, and U Mansmann. Graphical models illustrated complex associations between variables describing human functioning. *J Clin Epidemiol*, (62):922–33, 2009.

[21] R Strobl, E Grill, and U Mansmann. Graphical modeling of binary data using the LASSO: a simulation study. *BMC Med Res Methodol*, (12):16, 2012.

[22] JD Reinhardt, BA Fellinghauer, R Strobl, and Stucki G. Dimension reduction in human functioning and disability outcomes research: graphical models versus principal components analysis. *Disability and rehabilitation*, (32):1000–10, 2010.

[23] S. Becker, R. Strobl, A. Cieza, E. Grill, U. Harreus, and U. Tschiesner. Graphical modeling can be used to illustrate associations between variables describing functioning in head and neck cancer patients. *J Clin Epidemiol*, (64):885–92, 2011.

KAPITEL 2

GRAPHICAL MODELS ILLUSTRATED COMPLEX ASSOCIATIONS BETWEEN VARIABLES DESCRIBING HUMAN FUNCTIONING

Ralf Strobl[a], Gerold Stucki[a,b,c], Eva Grill[a], Martin Müller[a], Ulrich Mansmann[d]

[a]*Institute for Health and Rehabilitation Sciences, Ludwig-Maximilians-Universität München, Munich, Germany*

[b]*Department of Physical and Rehabilitation Medicine, University Hospital Munich, Ludwig-Maximilians-Universität München, Munich, Germany*

[c]*Swiss Paraplegic Research, Nottwil, Switzerland*

[d]*Department of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany*

This Chapter was published as

Strobl R, Stucki G, Grill E, Muller M, Mansmann U: **Graphical models illustrated complex associations between variables describing human functioning.** *J Clin Epidemiol* 2009, **62**:922-933.

Abstract

Objective: To examine whether graphical modeling is a potentially useful method for the study of human functioning using data collected by means of the International Classification of Functioning, Disability and Health (ICF).

Study Design and Setting: The applicability of the method was examined in a convenience sample of 616 patients from a crosssectional multicentric study undergoing early postacute rehabilitation. Functioning was qualified using 115 second-level ICF categories. The modeling was carried out on a data set with imputed missing values. The least absolute shrinkage and selection operator (LASSO) for generalized linear models was used to identify conditional dependencies between the ICF categories. Bootstrap aggregating was used to enhance the accuracy and validity of model selection.

Results: The resulting graph showed highly meaningful relationships. For example, one structure centered around speaking and included three paths addressing conversation, speech functions, and mental functions of language.

Conclusion: Graphical modeling of human functioning using data collected by means of the ICF yields clinically meaningful results. The structures found may be the basis for the identification of suitable targets for rehabilitation interventions, the identification of confounders and intermediate variables, and the selection of parsimonious sets of variables for multivariate epidemiological modeling.

Keywords: *Graphical models; ICF; Functioning; Rehabilitation; Statistical models; Classification*

2.1 INTRODUCTION

Functioning and disability are universal human experiences [1, 2]. Limitations in functioning occur in relation to acute or chronic health conditions, or aging. Modern societies aim to optimize individuals' functioning by applying curative, rehabilitative, and supportive health strategies [3].

Rehabilitation aims to optimize a person's capacity to interact with the environment by building on and strengthening the available resources and creating a facilitating environment [3]. To succeed, the relation between functioning and health conditions, and personal and environmental factors must be understood [4, 5].

There is a need to systematically study the distribution and determinants of functioning and the interactions between its components [4, 6, 7]. This relies on a conceptual framework and classification of the elements of functioning. The International Classification of Functioning, Disability and Health (ICF) [8] has been introduced and approved in 2001 as a suitable conceptual framework and classification.

Based on the ICF as a common language, it is possible to analyze functioning beyond the study of the incidence and prevalence of health conditions and beyond the limited aspects of functioning, such as activities of daily living [9], typically addressed with health-status measures [10, 11]. The current approach using health-status measures or clinical indices integrates many items into subscales or profiles, and hence, into few variables to address the methodological problems involved in multivariable and multivariate modeling. The use of fewer variables to avoid collinearity is at the expense of the ability to capture the underlying mechanisms [12]. As a consequence, distinct aspects of functioning may be masked.

With the ICF, it is possible to analyze the relationship of elements of functioning on the level of single categories. Because of the mentioned problems with multivariable and high-dimensional modeling, more appropriate methods need to be considered when examining the relationship of ICF categories and their potential applicability for the analysis of functioning data. One approach might be to adapt known methods established in other areas rather than to develop new methods.

A promising approach to describe the complex relationships in human functioning is graphical modeling. Graphical models form a comprehensive probabilistic tool to analyze and visualize dependencies between random variables.

In recent years, there have been numerous efforts to use graphical models in genetics, either for continuous or for dichotomous data. When analyzing genetic data, the large number of variables come along with fewer observations. The standard approach for dichotomous data would be to identify the relevant associations by using log linear models [13]. In the case of many variables, this exceeds the normal scope for log linear models, as all variables together with their interaction terms have to be included into the model. This results in a highly overparameterized model and unstable estimation of the associations [14]. Therefore, appropriate methods were recently developed to guide the model-building process in categorical data [15, 16].

Based on the experience in genetics and the structural similarity of genetic data with the ICF, we hypothesized that graphical modeling can be applied to the study of functioning data.

The objective of this study was to explore whether complex associations in human functioning can be properly analyzed using graphical-modeling techniques. If the method demonstrates known or established relationships between aspects of functioning, it will provide additional support for its face validity when analyzing this type of health data.

2.2 Methods

2.2.1 Principles of graphical models

Graphical models allow displaying the net-like dependence structures by drawing a graph describing the conditional dependencies between the variables. Thus, a graphical model is a graph with nodes representing random variables and edges between two nodes representing conditional dependency. Figure 2.1 shows an example of a simple graphical model. It shows the dependency structure of three variables (ICF categories) describing aspects of human functioning: the ability *to move around using equipment* (d465), the ability to *drink* (d560), and the ability to *eat* (d550).

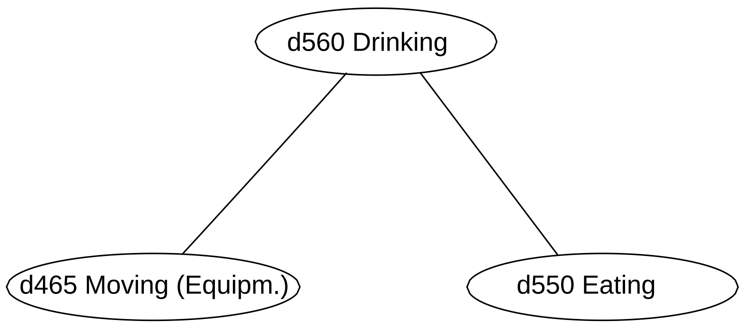

Figure 2.1 – *A simple graphical model.*

To examine conditional dependencies, it is important to understand the concept of conditional independence. In the example, d465 and d550 are considered conditionally independent given d560, if $f(d465|d550, d560) = f(d465|d560)$. Thus, information about restriction in eating does not give any additional information about restriction in moving with equipment, once we know about restriction in drinking. There will be no direct edge between two variables if they are conditionally independent given the remaining variables of the model. To give an example, in Figure 2.1, d465 and d550 are not connected by an edge, meaning they are conditionally independent given d560. The variable d560 can be seen as a confounder variable regarding the relation of d465 and d550. Thus, to study the relationship of d465 and d550, we should control for d560. However, if two variables are conditionally independent, it cannot be inferred that they are also marginally independent. Marginal independence ignores the possible effects of other variables, thus: $f(d465|d560) = f(d465)$. Both independencies have to be examined separately [17].

The edges used in graphical models to connect two variables indicate that the two variables are dependent even if controlling for other variables. Hence, when interpreting graphical models, one relies on edges, which indicate or represent associations between two variables based on the rejection of conditional independence.

To give an example, consider the association between *moving with equipment* (d465) and *eating* (d560) in Table 2.1. The variables are marginally associated. The odds ratio (OR) is 9.3, indicating that individuals with a mobility restriction have a higher probability for problems with eating. However, there is an association between mobility restriction and problems with drinking and between eating and drinking, indicating that the variable drinking confounds the association between moving and eating. Table 2.2 shows that the OR decreases when controlled, that is, stratified for drinking, resulting in OR = 1.3 in the stratum of individuals not restricted in drinking, and OR = 2.8 in the stratum of individuals restricted in drinking. Thus, the strength of the association shrinks conditionally on drinking. Figure 2.1 is a graphical expression of this relation. This phenomenon can go to the extent of reversal of the direction of association and is then known as Simpson's paradox [18].

Table 2.1 – *Contingency table describing the marginal relationship between* moving with equipment *(d465) and* eating *(d550)*

ICF category	d550-	d550+	Total
d465-	274	59	333
d465+	94	189	283
Total	368	248	616

\+ Indicates no restriction and - indicates restriction.

2.2.2 Study design and database

The method was applied to data from a multicentric cross-sectional survey in patients undergoing early postacute rehabilitation [19]. Patients were recruited from five early postacute rehabilitation units between June 2002 and April 2003. Of these, two units are specialized in neurological conditions, one is specialized in elderly patients, and two units located in acute hospital care for patients across a wide range of diagnosis.

Methods of data collection have been described in detail [19, 20, 21]. In brief, patients were interviewed by health professionals trained in the application and

Table 2.2 – *Contingency table describing the marginal relationship between* moving with equipment *(d465) and* eating *(d550) stratified by* drinking *(d560)*

ICF category	d550-	d550+	Total
d560-			
d465-	227	5	232
d465+	36	1	37
Total	263	6	269
d560+			
d465-	47	54	101
d465+	58	188	246
Total	105	242	347

+ Indicates no restriction and - indicates restriction.

principles of the ICF. Information on patients' functioning was acquired by personal interview or, if information was not obtainable from the patient, health professionals of the corresponding wards or relatives of the patient were asked. A positive vote from the ethics committee was obtained before start. Informed consent was obtained from patients, or, if a patient was unable to make an informed decision, from the patient's care giver.

2.2.3 Overview on International Classification of Functioning, Disability and Health coding

The ICF can be divided into two parts. The first part covers functioning and disability and includes the components Body Functions (b), Body Structures (s), and Activities and Participation (d), and the second one covers the contextual factors, namely Environmental and Personal Factors. Body Functions and Body Structures describe physiological functions and anatomical parts. Activities and Participation refer to task execution and the involvement of individuals in real-life situations. In the ICF classification, the letters b, s, and d are followed by a numeric code starting with the chapter number (one digit) followed by the second (two digits), third, and fourth levels (one digit each).

For this study, we restricted the analysis to the functioning and disability part of the ICF. The second-level categories of the ICF components Body Functions, Body Structures, and Activities and Participation were used. Each second-level category of the components b, s, and d was graded with the qualifier 0 for 'no impairment/limitation' and 1 for 'impairment/limitation'. A total of 115 second-level categories of the ICF were investigated in this study.

2.2.4 Model building

General approach

Figure 2.2 gives a flow chart of the different methods involved in the model-development process. The specific methods are explicitly explained in the following sections. We first imputed the missing values to be able to work with complete observations. Then, we carried out bootstrap replications to enhance the accuracy and validity of model selection. The modeling was finally done on the imputed and replicated data sets with the final result being the composite result over all data sets.

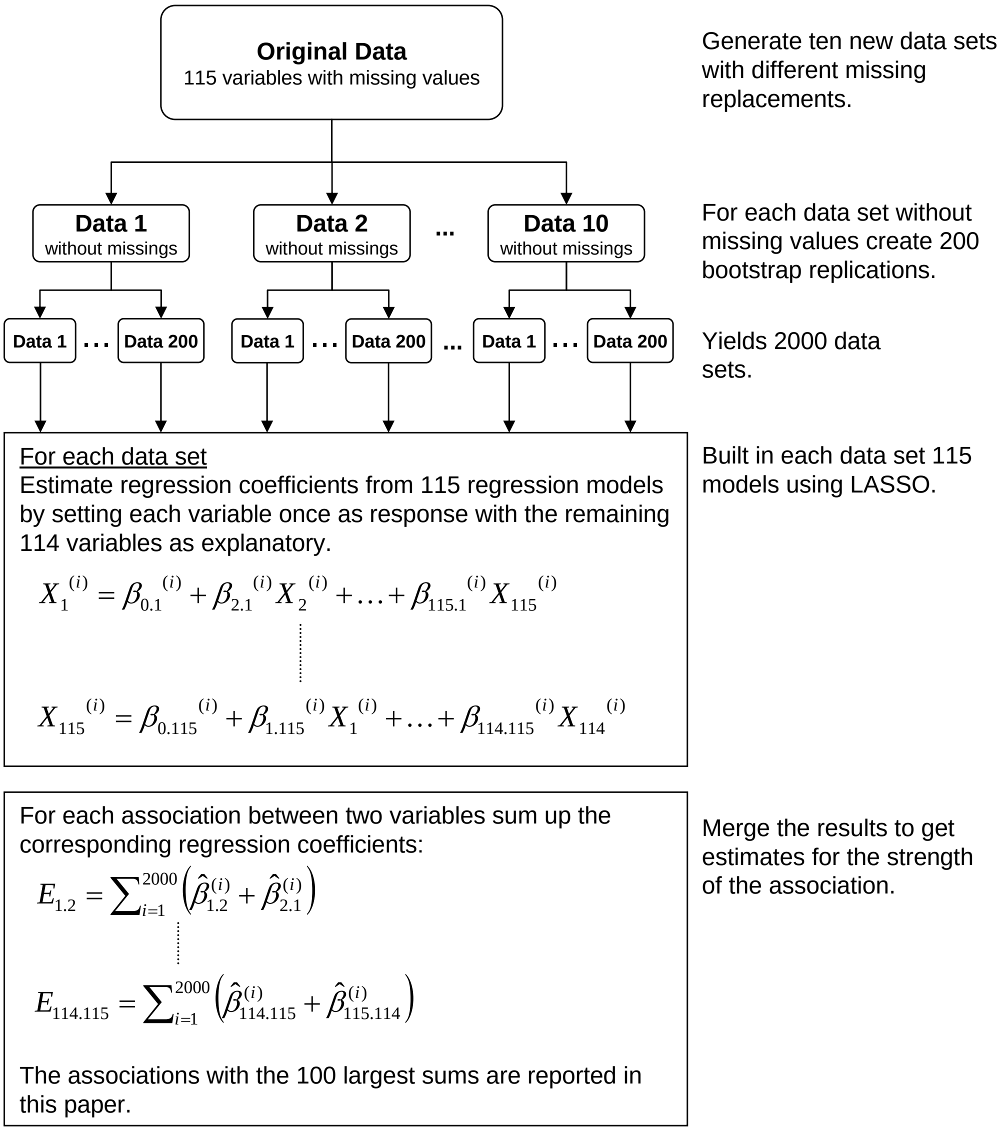

Figure 2.2 – *Flow chart of the different methods involved in the model-development process.*

Handling of missing data

Missing values may introduce bias and lead to false conclusions under certain circumstances. Still, appropriate handling and reporting of missing data is a common problem [22]. To handle the missing data in our data set, we used multiple imputations [23]. This method was shown to be superior regarding variance in simulation and real data settings to complete case analysis and showed similar results to other common missing-value approaches [24, 25]. Multiple imputations generate m versions of the original data set, with varying missing-value replacements in each version. Simulation studies demonstrated that even with five generated data sets, multiple imputations yield valid results [25]. Our choice of $m = 10$ is well above it. We specified a conditional distribution for each incomplete variable by a logistic regression using information from all other variables. By iterative Gibbs sampling, 10 complete data sets were drawn from this multivariate distribution [26]. Each of these 10 imputed data sets was then analyzed using common complete case methods. Thus, for each of the imputed data sets, a regression model was estimated and the results were averaged to get a final single-point estimate [27]. For imputation, we used the MICE (Multivariate Imputation by Chained Equations) package running under R [28, 29].

Enhancing performance

Graphical models are susceptible to small changes in the data set leading to unstable results. A common method to enhance unstable procedures is to use bootstrap aggregating (bagging) [30]. Bagging produces several models based on bootstrap replicates of the original data set. The multiple versions are then aggregated. Bootstrap aggregating can stabilize the outcome of a model and enhance accuracy [31, 32]. Additionally, this way, the performance exceeds that of a single model [33]. Since majority vote is a standard way to aggregate multiple models we used weighted majority vote with coefficients as weights. This yielded the final estimate of the strength of the associations.

Typically, at least 50 replications are generated [30]. We generated 200 bootstrap replications to obtain further precise and valid result. These 200 replications of the 10 imputed data sets generated 2000 data sets for the final analysis. Bagging was run under R using a self-constructed code.

Variable-selection procedure

In the usual regression situation, the model is built by minimizing the residual squared error. If there are many independent variables to be considered, ordinary least square estimates are not biased but their variance can be very high. To analyze the high-dimensional data set, we adapted a newly established method based on the least absolute shrinkage and selection operator (LASSO) [34]. The LASSO restricts the minimization of the residual sum of squares to an upper bound t (the penalty value) for the sum of the absolute values of the coefficients. This reduces the variance of the estimate and also acts as a variable-selection procedure by setting unimportant coefficients to zero. The approach can be formalized as:

$$\min_{\beta_1 \ldots \beta_p} \frac{1}{2} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2$$

subject to $\sum_{j=1}^{p} |\beta_j| \leq t$ for some $t > 0$.

This method was used to discover the conditional independencies in the neighborhood of a variable. The neighborhood of variable a (ne_a) is defined as the set of conditionally dependent variables of a. Thus, given all variables in ne_a, a is conditionally independent of all remaining variables. Usually, covariance-selection methods are used to sort out conditionally independent variables [35]. Covariance selection aims to identify the zero elements in the inverse of the covariance matrix. Meinshausen and Buehlmann [15] showed that neighborhood selection by means of the LASSO [34] is superior to common covariance- selection methods, in particular, if the number of variables exceeds the number of observations. They defined a neighborhood ne_a as the set of predictor variables corresponding to nonzero coefficients in a prediction model by estimating the conditional independence separately for each variable. To be able to use this approach for binary data, we modified it by using the LASSO for logistic regression [36]. The following steps led to the final graph:

1. Estimate the coefficients in a LASSO-style logistic regression using each variable as an outcome variable and the remainder as predictors.

2. Identify the neighborhood of variable a (ne_a) as the set of variables corresponding to nonzero coefficients: $ne_a = \left\{ b \in \nu : \hat{\beta}_{\nu}^{LASSO} \right\}$ with ν as the set of variables (nodes).

3. Define the set of conditional relationships (the edge set) E as:

$$E = \{(a, b)) \, | a \in ne_b \vee b \in ne_a\}$$

To get an impression if the method is capable of retrieving existing structures, we performed a simulation study. We generated binary random data sets corresponding to known directed dependency structures. Data were then modeled accordingly. The resulting graphs were compared with the inbuilt known structure. The method could retrieve the main inbuilt structures. Bagging improved the accuracy of the method. The simulation further showed that the choice of the penalty parameter in the LASSO setting was not essential. However, a penalty t value of 0.5 was slightly superior compared with either 0.3, 0.7, 1.0, or 1.5. Thus, we used a penalty value of 0.5 yielding a mean neighborhood size of eight variables. The LASSO regression was performed using the `lasso2` package running under R.

2.2.5 Representation of the graphs

The visualization of the graph is done using the R package Rgraphviz [37] using an automatic algorithm for drawing undirected graphs [38]. An edge in the graph represents the partial correlation between two nodes. To give an example, the chain $A - B - C$ would result from a situation where A is associated with B, and B is associated with C. There is no association between A and C once B is accounted for. If the associations $A - B$ and $B - C$ do not fully explain $A - C$, a triangle $A - B - C$ would be the result.

In an undirected graph, there is no core or starting node. The algorithm randomly chooses a variable as starting node. Model selection continues until the optimal, most parsimonious configuration of the graph is found.

We added the ORs to the edges of our graphs estimated from the penalized logistic regression models using $OR = \exp(\text{regression coefficient})$. This allows the derivation of local logistic regression models from the neighborhood of a node formalizing the distribution of the node's values. Applying Gibbs sampling techniques would recover the complete multivariate binary distribution from the local logistic regressions. In favor of readability, we did not add the baseline odds for the local regression to the respective nodes in the graphs.

The thickness of the edges corresponds to the strength of the association using the following rule: the 20% with the strongest associations are printed in bold and the 20% with the weakest associations are in gray shade. Negative associations are plotted with dashed lines. The 20% strongest and weakest associations are identified once in the entire model as opposed to the substructures.

2.3 Results

Six hundred and sixteen patients were included in the analyses, with 46.0% being males. Their mean age was 63.2 years, ranging from 18.0 to 98.0 years. 12.8% of the patients had been admitted because of a cardiopulmonary problem, 30.4% had a musculoskeletal problem, and 55.7% had a neurological problem. We could not clearly allocate 1% of the patients to a certain diagnosis groups. There were only six complete cases without missing values.

The final model was based on all the variables introduced in Table 2.3, together with gender and age. The ICF coding used in the graphs is described in Table 2.3. The baseline odds of the respective ICF categories are presented in the supplemental material. Figure 2.3 shows the 100 strongest associations of the resulting graph. The slope of the estimated strength of the first 300 associations is presented in the supplemental material. Different large structures could be found in this model. We display the two structures with the longest paths and the overall largest number of variables in Figures 2.4 and 2.5 and describe them separately. A short description

Table 2.3 – *ICF - short description of the ICF categories used for the graphs*

ICF	Description	ICF	Description
b114	Orientation functions	d110	Watching
b126	Temperament and personality functions	d115	Listening
b130	Energy and drive functions	d120	Other purposeful sensing
b140	Attention functions	d130	Copying
b144	Memory functions	d135	Rehearsing
b147	Psychomotor functions	d155	Acquiring skills
b152	Emotional functions	d160	Focusing attention
b160	Thought functions	d166	Reading
b164	Higher level cognitive functions	d170	Writing
b167	Mental functions of language	d175	Solving problems
b176	Mental functions of sequencing complex movements	d177	Making decisions
b180	Experience of self and time functions	d230	Carrying out daily routine
b210	Seeing functions	d240	Handling stress and other psychological demands
b215	Functions of structures adjoining the eye	d310	Communicating with spoken messages
b235	Vestibular functions	d315	Communicating with nonspoken messages
b240	Sensations associated with hearing and vestibular functions	d330	Speaking
b260	Proprioceptive function	d335	Producing nonverbal messages
b265	Touch function	d350	Conversation
b270	Sensory functions related to temperature and other stimuli	d360	Using communication devices and techniques
b280	Sensation of pain	d410	Changing basic body position
b310	Voice functions	d415	Maintaining a body position
b320	Articulation functions	d420	Transferring oneself
b340	Alternative vocalization functions	d430	Lifting and carrying objects
b410	Heart functions	d440	Fine hand use
b415	Blood vessel functions	d445	Hand and arm use
b420	Blood pressure functions	d450	Walking
b440	Respiration functions	d460	Moving around in different locations
b445	Respiratory muscle functions	d465	Moving around using equipment
b450	Additional respiratory functions	d510	Washing oneself
b455	Exercise tolerance functions	d520	Caring for body parts
b460	Sensations associated with cardiovascular and respiratory functions	d530	Toileting
b510	Ingestion functions	d540	Dressing
b515	Digestive functions	d550	Eating
b525	Defecation functions	d560	Drinking
b535	Sensations associated with digestive functions	d570	Looking after one's health
b540	General metabolic functions	d760	Family relationships
b545	Water, mineral, and electrolyte balance functions	d940	Human rights
b550	Thermoregulatory functions	s110	Structure of brain
b610	Urinary excretory functions	s120	Spinal core and related structures
b620	Urination functions	s130	Structure of meninges
b630	Sensations associated with urinary functions	s410	Structure of cardiovascular system
b710	Mobility of joint functions	s430	Structure of respiratory system
b715	Stability of joint functions	s530	Structure of stomach
b730	Muscle power functions	s710	Structure of head and neck region
b735	Muscle tone functions	s720	Structure of shoulder region
b740	Muscle endurance functions	s730	Structure of upper extremity
b755	Involuntary movement reaction functions	s740	Structure of pelvic region
b760	Control of voluntary movement functions	s750	Structure of lower extremity
b770	Gait pattern functions	s760	Structure of trunk
b810	Protective functions of the skin	s810	Structure of areas of skin
b820	Repair functions of the skin	s840	Structure of hair

Abbreviation: ICF, International Classification of Functioning, Disability and Health.

of the ICF categories has been added to the node labels. As there are no negative associations, there are no dashed lines.

The structure in Fig. 2.4 shows categories set around the activity category *speaking* (d330). The categories are aggregated into three different paths. The first path describes the ability to have a *conversation* (d350). The category *conversation* is

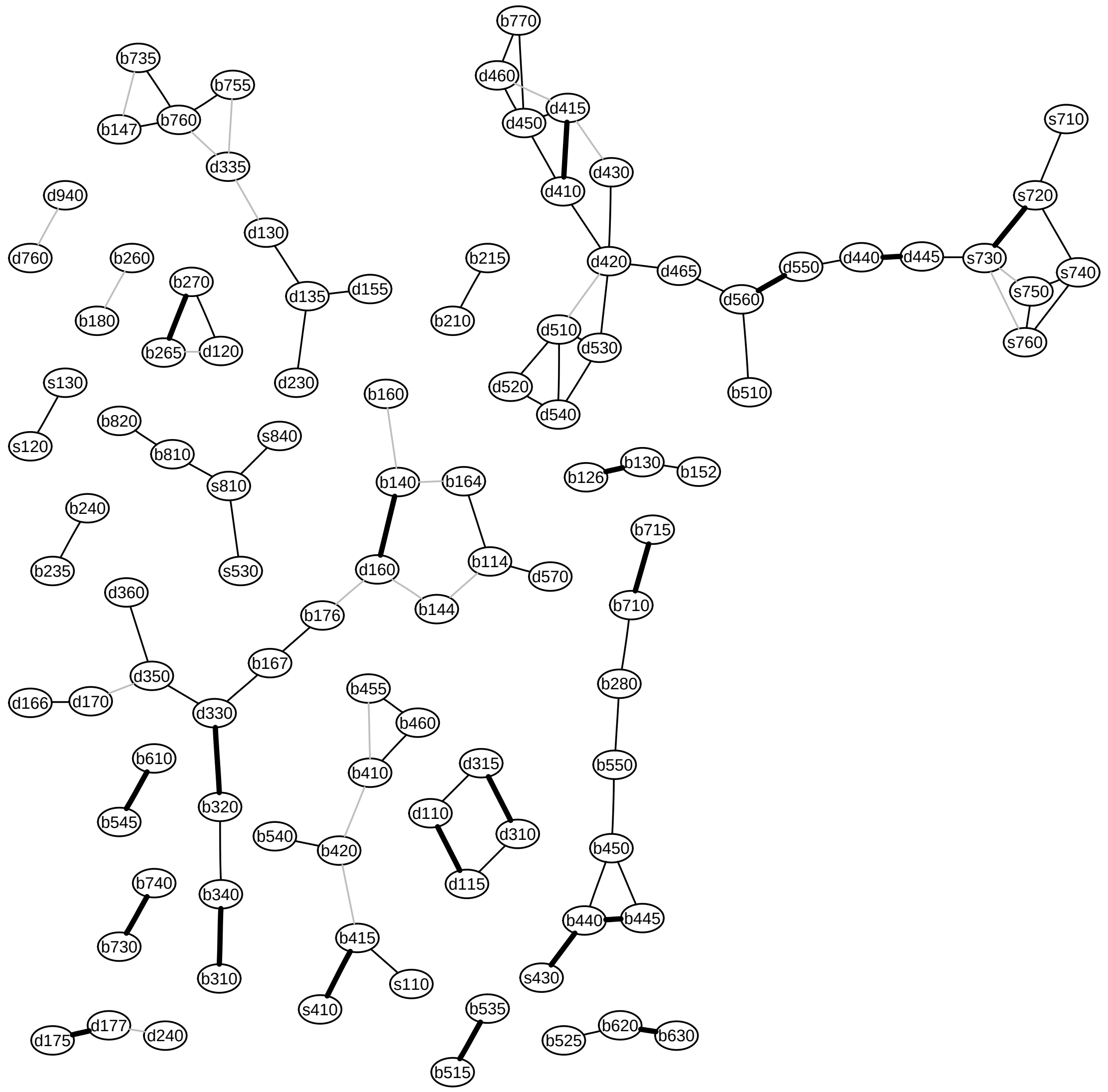

Figure 2.3 – *The graphical model with the 100 strongest associations. The presentation of the edges corresponds to the strength of association, that is, the 20 edges with the strongest association are shown in bold and the 20 edges with the weakest association are shown in gray shade.*

itself the center of the categories *speaking* and a structure with the abilities of *using communication devices* (d360) and *reading* and *writing* (d170, d166). The second path centers on voice and speech functions, namely *articulation functions* (b320), *alternative vocalization functions* (b340), and *voice functions* (b310). The third path revolves around *mental functions of language* and of *sequencing movements* (b167, b176), followed by the ability to *focus attention* (d160). The path ends in a circle containing other aspects of mental functions, that is, *memory functions* (b144), *orientation functions* (b114), *higher level cognitive functions* (b164), and *attention functions* (b140). *Thought functions* (b160) is associated with *attention functions.* The ability to *look after one's health* (d570) is associated with *orientation functions.*

The left end of the structure in Fig. 2.5 includes two paths centered on *transfer-*

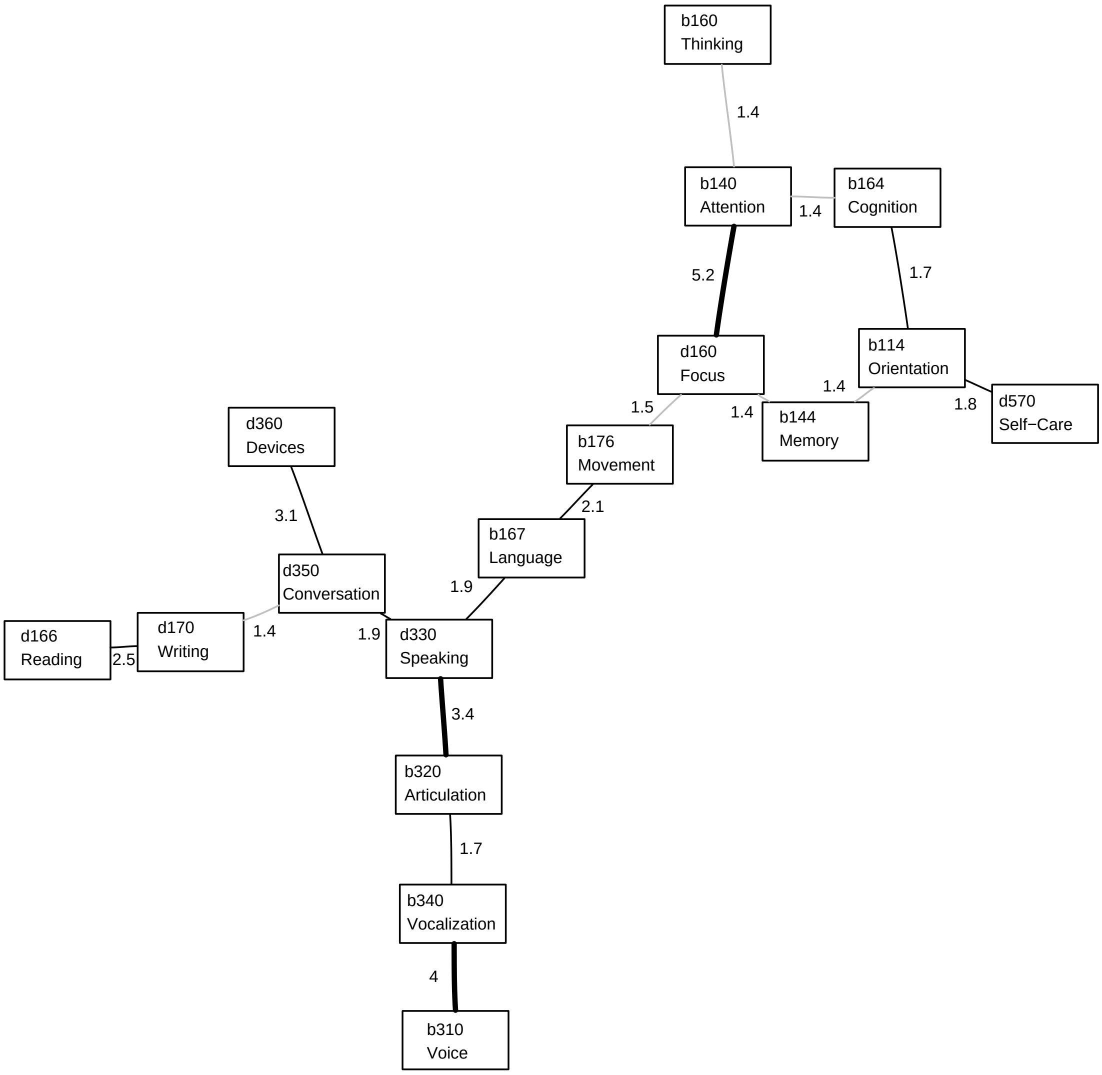

Figure 2.4 – *A structure revolving around* speaking *(d330). The presentation of the edges corresponds to the strength of association, that is the 20 edges with the strongest associations are shown in bold and the 20 edges with the weakest associations are shown in gray shade. Each edge is labeled with the odds ratio (OR) of the association between the two variables estimated from the penalized logistic regression models using OR = exp(regression coefficient). To put this into context, higher ORs indicate stronger positive associations. An OR of 1 would indicate no association and an OR smaller than 1 would indicate a negative association.*

ring oneself (d420). One path is built on the different aspects of self-care *toileting* (d530), *dressing* (d540), *washing oneself* (d510), and *caring for body parts* (d520), and the other one is characterized by basic aspects of mobility, such as *maintaining* (d415) and *changing* (d410) *body positions*, *lifting and carrying objects* (d430), *moving around in different locations* (d460), and *walking* (d450) together with the Body Function *gait pattern* (b770). *Moving around with the help of equipment* (d465) is the start of the third path associated with *transferring oneself* (d420) and further

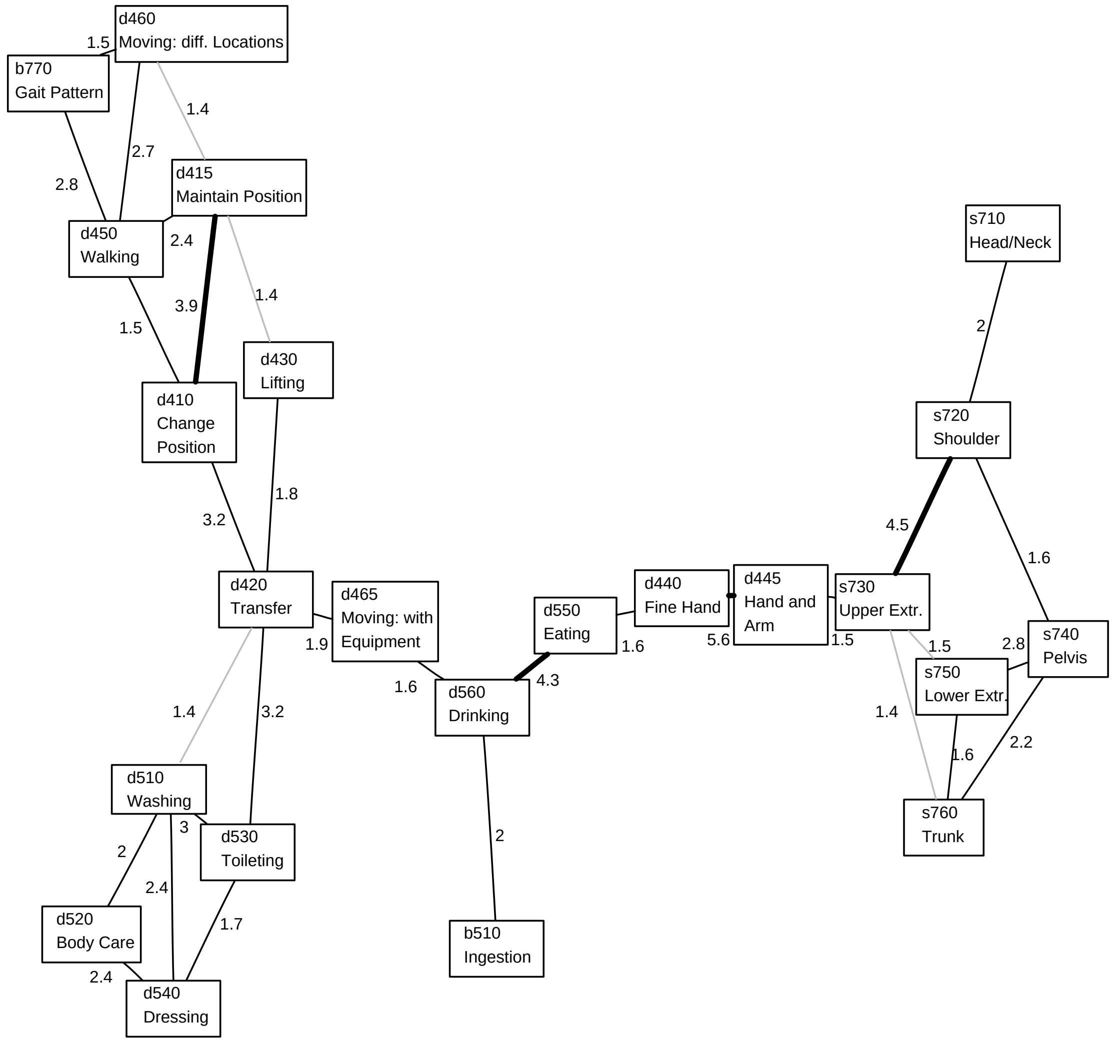

Figure 2.5 – *A structure resolving around* transferring oneself *(d420). See Fig. 2.4 for details.*

associated with *drinking* (d560) and *eating* (d550). *Drinking* is also associated with *ingestion functions* (b510) and *eating* with *fine hand use* (d440) and *hand and arm use* (d445). *Hand and arm use* is associated with a cluster of categories belonging to different body structures related to movement (s710 - s760).

2.4 Discussion

This study introduces graphical models to examine the complex structures of human functioning based on the example of patients in need of early postacute rehabilitation. We found that graphical models can be applied to and provide insights into the relationship between different aspects of functioning as represented by the ICF categories. The results have to be seen both in the light of their clinical and methodological implications.

2.4.1 Issues of clinical validity

Many of the associations found are clinically meaningful and give evidence of the high face validity of the method. The structure in Fig. 2.4 may serve as a case in point. In this structure, three paths were set around the category *speaking* (d330) relating to three different aspects of speech and language disorders, namely 'voice aspects', 'mental aspects', and 'aspects of producing communication'. One path, 'voice aspects', containing articulation and voice functions (b320, b340, b310) represents problems with expressing speech sounds, such as dysarthria or speech apraxia [39, 40]. A second path, 'mental aspects', represents mental functions associated with language: *mental functions of language* (b167), *memory* (b144), and *attention* (b140). Impairments of these functions are well known from patients with brainstem lesions [41]. Additionally, *thought functions* (b160) are part of this path. It has been controversially discussed whether aphasic impairments are associated with disorders of thinking. The discussion is ongoing at least since 1870, when aphasia was proposed to be an asymbolia [42] - posing questions, such as do we need words to think?; Is someone with a language disorder impaired in thinking? The matter is not resolved until today [43, 44]. Finally, a third path, 'aspects of producing communication', contains the communicative aspects of language, namely *conversation* (d350), which relate not only to speaking but also to *using communication devices* (d360) and *reading* and *writing* (d170, d166). Other aspects of receiving language, such as *understanding spoken language* (d310), which are also components of communication, are represented by another structure, namely connecting the categories *receiving nonverbal messages* (d315), *watching* (d110), and *listening* (d115).

Our example illustrates the complex relationships of variables within the framework of the ICF components. The discussed structure around *speaking* (d330) may also serve as a case in point for structures relating Body Functions with Activities and Participation. Few associations of other structures seem counterintuitive at least at first glance. To give an example, the association of *structures of stomach* (s530) and *structures of skin* (s810) seems counterintuitive. This can be explained by a rather large subset of patients receiving percutaneous enteral nutritional support. Percutaneous endoscopic gastrostomy would be needed to place the feeding device, involving lesions of stomach and skin.

2.4.2 Methodological considerations and applications

The results are equally meaningful from a methodological perspective, namely regarding the identification of confounders, intermediate variables, potential constructs and scales, subsets for regression analysis, and intervention targets.

As for the issue of confounders and intermediate variables, when analyzing the effect of one variable on another, potentially confounding variables may distort or even reverse the true association. Mathematical solutions to cope with confounding depend on the number of covariates involved and are not always obvious [45]. Graphical models not only give an idea about potential confounders, but also differentiate them from intermediate variables. To give an example, body structures related to movement (s710-s760) are associated with the corresponding activities, such as *hand and arm use* (d445) (Fig. 2.5). These activities are associated with the ability to eat and drink (d550, d560), but eating and drinking are also associated with *ingestion functions* (b510). The graphical model shows that eating and drinking are intermediate variables in the pathway, not confounders. In contrast, in the same structure (Fig. 2.5), the three categories, *changing body position* (d410), *maintaining body position* (d415), and *walking* (d450), form a triangle, meaning each category is independently associated with the other two. Likewise, graphical models provide an easy solution for which variable to control when several interrelated variables qualify as potential confounders [17].

When constructing scales for clinical measurement, variables with a common underlying construct are valid candidates. Circular paths illustrate the high interdependency of a number of categories, such as the pentagon-type structure with *memory functions* (b144), *orientation functions* (b114), *cognitive functions* (b164), *attention functions* (b140), and *focusing attention* (d160). Theoretically, circular paths in graphical models contain variables which are sufficiently correlated to form a common dimension for measurement. They may, therefore, be candidates for inclusion to a scale [46].

As to regression modeling, graphical models can be used to simplify further studies and to construct more parsimonious but equally valid models. To give an example, it would be sufficient in a study about factors influencing *speaking* (d330) to just look at the direct neighbors of speaking as shown in the final graph, namely *conversation* (d350), *mental functions of language* (b167), and *articulation functions* (b320). Possibly, hierarchical models can be used to describe more complex relationships.

As to clinical interventions, graphical models can be used to identify targets for therapy. By targeting central categories in graphs, other categories will be influenced. To give an example, in our study, the category *transferring oneself* (d420) is directly associated with several categories of the ICF chapter self-care (d510-d540). One could easily assume that any improvement in transfer abilities would translate into improved ability to *wash* (d510), *care for body parts* (d520), *toilet* (d530), and *dress* (d540).

In general, interesting geometrical objects can give indications how to analyze data. To give an example, the 'bowtie'- like structure (Fig. 2.3) connects *muscle tone functions* (b735), *psychomotor functions* (b147), *control of voluntary movement* (b760), *involuntary movement reaction* (b755), and *producing nonverbal messages* (d335). Here, b760 in the middle of the structure could serve as a proxy for both b147 and b735. For example, when investigating the association between d355 and b755, it would be sufficient to adjust for b760.

2.4.3 Comparison with other multivariate exploratory methods

The method used in this study bears analogy to other multivariate exploratory methods, namely factor analysis, path analysis, and Bayesian networks. Although the aims of these methods help researchers understand the structure of a high-dimensional distribution, they focus on different aspects.

Factor analysis defines a new coordinate system of relevant dimensions and recalculates the coordinates of observations as loadings. It aims to reduce the number of manifest (i.e., existing) variables by constructing a smaller set of latent variables. In contrast, graphical modeling as applied in this article does not aim to reduce dimensions but to provide an interpretable representation of a complex correlation structure.

Path analysis, or structural equation modeling (SEM), studies directed causal associations between variables [47] and enriched correlation structure by causal meaning [17]. They aim to confirm a preset causal model. In contrast, our application is exploratory with respect to interaction structures, but also allows confirmatory assessment of hypothesized associations. With regard to our application of graphical models to human functioning, future investigations will, for example, test putative intervention effects by comparing changes of interaction structures over time.

Bayesian networks or directed acyclic graphs (DAG), as an example for hierarchical structuring, decompose multivariate distributions in directed layers of dependent simple distributions. Our objective is to elucidate dependence structure without the ambition of determining the direction of relationships.

2.4.4 Limitations

Firstly, the model-building process has to be critically addressed. The chosen imputation method may induce artificial associations. Multiple imputations, however, have been shown to be superior to other methods, such as imputation by the mean, and to yield valid results [48, 49, 50]. Bootstrap aggregating was used to improve the model, irrespective of the penalty value for the LASSO regression. The simulation study showed that the critical choice of the penalty value might be avoided. However, it is still to be investigated if the validity of the choice of critical parameters can be improved.

Secondly, we have used graphical modeling as a tool for exploratory data analysis without any preset hypothesis. Because one may rather have the goal to quantify direction and strength of effects than to illustrate structures, there is also the need to apply methods to test and to confirm hypotheses about graphs.

Thirdly, the data-collection procedure may have been biased. The interviewers may have introduced artificial dependencies by their own interpretation of the associations. However, we are sufficiently confident that the standardization of the procedure including interviewer training and supervision should have ruled this out [19, 20, 21].

Fourthly, our interpretation of results may have been overly optimistic. Further studies are needed to assess the stability and reproducibility of the results. There is no evidence of the post hoc reliability of graph interpretation. It would be interesting to see if clinical experts were able to distinguish between randomly generated graphs and graphs originating from real world data.

Fifthly, we did not differentiate between specific impairments and limitations owing to different health conditions, because we wanted to investigate functioning within the early postacute situation. Further research is needed to specifically address problems arising in, for example, stroke patients in contrast to other conditions.

2.5 Conclusion

Graphical modeling of functioning using data collected with the ICF yields clinically meaningful results. It is, thus, a promising method to gain insight into the complex association structure of human functioning. It can be the basis for the identification of confounders and intermediate variables, potential constructs, scales and indices, subsets for regression model development, and intervention targets.

Acknowledgments

This project was supported by the German Ministry of Health and Social Security (BMGS) grant no 124-43164- 1/501 and by the LMUinnovativ project Münchner Zentrum für Gesundheitswissenschaften (TP 1 and TP 2).

We thank the teams at the Neurological Therapy Centre Burgau, at the Neurological Hospital Bad Aibling, at the Geriatric Rehabilitation Hospital Wurzburg, at the Department of Physical Medicine and Rehabilitation, Hospital Munich-Bogenhausen, and at the Department of Physical Medicine and Rehabilitation at the University Hospital Munich for data collection and support. We would like to thank Joe Ilmberger for valuable insights, two referees for many helpful recommendations, and Markus Kalisch for his comments on a later version of the manuscript.

Appendix

Supplementary material

Supplementary material can be found, in the online version, at 10.1016/j.jclinepi.2009.01.018.

Literaturverzeichnis

[1] J. E. Bickenbach, S. Chatterji, E. M. Badley, and T. B. Ustun. Models of disablement, universalism and the international classification of impairments, disabilities and handicaps. *Soc Sci Med*, 48(9):1173–87, 1999.

[2] I. K. Zola. Toward the necessary universalizing of a disability policy. *Milbank Q*, 67 Suppl 2 Pt 2:401–28, 1989.

[3] G. Stucki, A. Cieza, and J. Melvin. The International Classification of Functioning, Disability and Health (ICF): a unifying model for the conceptual description of the rehabilitation strategy. *J Rehabil Med*, 39(4):279–85, 2007.

[4] G. Stucki and G. Grimby. Organizing human functioning and rehabilitation research into distinct scientific fields. Part I: Developing a comprehensive structure from the cell to society. *J Rehabil Med*, 39(4):293–8, 2007.

[5] G. Stucki, J. D. Reinhardt, and G. Grimby. Organizing human functioning and rehabilitation research into distinct scientific fields. Part II: Conceptual descriptions and domains for research. *J Rehabil Med*, 39(4):299–307, 2007.

[6] 58th World Health Assembly. Resolution 114: disability, including prevention, management and rehabilitation. Adopted May 2005. Technical report, World Health Organization, 2005.

[7] Walter R. Frontera, Marcus J. Fuhrer, Alan M. Jette, Leighton Chan, Rory A. Cooper, Pamela W. Duncan, John D. Kemp, Kenneth J. Ottenbacher, P Hunter Peckham, Elliot J. Roth, and Denise G. Tate. Building research capacity. Proceedings of the 2005 Rehabilitation Medicine Summit. April 28-29, 2005. Washington, D.C., USA. *Am J Phys Med Rehabil*, 84(12):913–1052, 2005.

[8] WHO. International Classification of Functioning, Disability and Health: ICF, 2001.

[9] R. Hebert, C. Brayne, and D. Spiegelhalter. Incidence of functional decline and improvement in a community-dwelling, very elderly population. *Am J Epidemiol*, 145(10):935–44, 1997.

[10] M. D. Patel, K. Tilling, E. Lawrence, A. G. Rudd, C. D. Wolfe, and C. McKevitt. Relationships between long-term stroke disability, handicap and health-related quality of life. *Age Ageing*, 35(3):273–9, 2006.

[11] P. Raina, S. Dukeshire, J. Lindsay, and L. W. Chambers. Chronic conditions and disabilities among seniors: an analysis of population-based health and activity limitation surveys. *Ann Epidemiol*, 8(6):402–9, 1998.

[12] G. Kwakkel, B. Kollen, and E. Lindeman. Understanding the pattern of functional recovery after stroke: facts and theories. *Restor Neurol Neurosci*, 22(3-5):281–99, 2004.

[13] Alan Agresti. *Categorical Data Analysis.* Wiley, New Jersey, 2002.

[14] Dirk Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, 19:2271–2282, 2003.

[15] Nicolai Meinshausen and Peter Buehlmann. High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*, 34:1436–1462, 2006.

[16] Juliane Schaefer and Korbinian Strimmer. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21:754–764, 2005.

[17] Judea Pearl. *Causality: Models. Reasoning and Inference.* Cambridge University Press, New York, 2000.

[18] E.H. Simpson. The interpretation of interaction in contingency tables. *Journal of Royal Statistical Society*, 13:238–41, 1951.

[19] Eva Grill, Berthold Lipp, Christine Boldt, Gerold Stucki, and Eberhard Koenig. Identification of relevant ICF categories by patients with neurological conditions in early post-acute rehabilitation facilities. *Disability and Rehabilitation*, 27:459–465, 2005.

[20] Eva Grill, M Harder, L Fischbacher, Christine Boldt, R Mittrach, and Gerold Stucki. Identification of relevant ICF categories by patients in early post-acute rehabilitation facilities. *Phys Med Rehab Kuror*, 15:168–173, 2005.

[21] E. Grill, G. Stucki, C. Boldt, S. Joisten, and W. Swoboda. Identification of relevant ICF categories by geriatric patients in an early post-acute rehabilitation facility. *Disabil Rehabil*, 27(7-8):467–73, 2005.

[22] A. Burton and D. G. Altman. Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. *Br J Cancer*, 91(1):4–8, 2004.

[23] D. B. Rubin. Multiple imputations in sample surveys - a phenomenological Baysian approach to nonresponse. *Proceedings of the Survey Research Method Sections, American Statistical Association*, pages 20–34, 1978.

[24] N. J. Horton and K. P. Kleinman. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *Am Stat*, 61(1):79–90, 2007.

[25] Joseph G. Ibrahim, Ming-Hui Chen, Stuart R. Lipsitz, and Amy H. Herring. Missing-Data Methods for Generalized Linear Models: A Comparative Review. *Journal of the American Statistical Association*, 100:332–346, 2005.

[26] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 2 edition, 2004.

[27] Roderick J Little and Donald B Rubin. *Statistical Analysis with Missing Data*. Wiley, New York, 2002.

[28] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.

[29] S. van Buuren and C.G.M. Oudshoorn. Multivariate imputation by chained equations: Mice V1.0 Users Manual, 2000.

[30] L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.

[31] Peter Buehlmann and Yu Bin. Analyzing bagging. *The Annals of Statistics*, 30:927–961, 2002.

[32] Andreas Buja and Werner Stuetzle. Observations on Bagging. *Statistica Sinica*, 16:323–352, 2006.

[33] David Opitz and Richard Maclin. Popular Ensemble Methods: An Empirical Study. *Journal of Artificial Intelligence Research*, 11:169–198, 1999.

[34] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58:267–288, 1996.

[35] A.P. Dempster. Covariance Selection. *Biometrics*, 28(1):157–175, 1972.

[36] Justin Lokhorst. The LASSO and Generalised Linear Models. *Honours Projects*, Department of Statistics. South Australia, Australia: The University of Adelaide; 1999.

[37] Jeff Gentry, Li Long, Robert Gentleman, and Seth Falcon. Rgraphviz: Provides plotting capabilities for R graph objects, 2008.

[38] Tomihisa Kamada and Satoru Kawai. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31:7–15, 1989.

[39] Joseph R. Duffy. *Motor speech disorders*. Mosby, Boston, 1995.

[40] R. D. Kent. Research on speech motor control and its disorders: a review and prospective. *J Commun Disord*, 33(5):391–427; quiz 428, 2000.

[41] Muriel Deutsch Lezak. *Neuropsychological Assessment.* Oxford University Press, New York, 2004.

[42] F. C. Finkelnburg. Referat über Aphasie am 21. März 1870 in der Medicinischen Section der Niederrheinischen Gesellschaft Bonn, 1870.

[43] Peter Bachman and Tyrone D Cannon. Cognitive and Neuroscience Aspects of Disorder. pages 493–528. Cambridge University Press, New York, 2005.

[44] Lila Gleitman and Anna Papafragou. Language and Thought. In *The Cambridge Handbook of Thinking and Reasoning*, pages 633–662. Cambridge University Press, New York, 2005.

[45] S. Greenland and H. Morgenstern. Confounding in health research. *Annu Rev Public Health*, 22:189–212, 2001.

[46] Kenneth A. Bollen. Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53(1):605–634, 2002.

[47] Sewall Wright. The method of path coefficient. *Annals of Mathematical Statistics*, 5:161–215, 1934.

[48] A. R. Donders, G. J. van der Heijden, T. Stijnen, and K. G. Moons. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*, 59(10):1087–91, 2006.

[49] F. M. Shrive, H. Stuart, H. Quan, and W. A. Ghali. Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. *BMC Med Res Methodol*, 6:57, 2006.

[50] G. J. van der Heijden, A. R. Donders, T. Stijnen, and K. G. Moons. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol*, 59(10):1102–9, 2006.

KAPITEL 3

GRAPHICAL MODELING OF BINARY DATA USING THE LASSO: A SIMULATION STUDY

Ralf Strobl[a], Eva Grill[a,b], Ulrich Mansmann[a]

[c] *Institute for Medical Informatics, Biometrics and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany*

[b] *Faculty of Health and Nursing Sciences, West Saxon University of Applied Sciences, Zwickau, Germany*

This Chapter was published as

Strobl R, Grill E, Mansmann U: **Graphical modeling of binary data using the LASSO: a simulation study.** *BMC Medical Research Methodology* 2012, **12**:16.

Abstract

Background: Graphical models were identified as a promising new approach to modeling high-dimensional clinical data. They provided a probabilistic tool to display, analyze and visualize the net-like dependence structures by drawing a graph describing the conditional dependencies between the variables. Until now, the main focus of research was on building Gaussian graphical models for continuous multivariate data following a multivariate normal distribution. Satisfactory solutions for binary data were missing. We adapted the method of Meinshausen and Bühlmann to binary data and used the LASSO for logistic regression. Objective of this paper was to examine the performance of the Bolasso to the development of graphical models for high dimensional binary data. We hypothesized that the performance of Bolasso is superior to competing LASSO methods to identify graphical models.

Methods: We analyzed the Bolasso to derive graphical models in comparison with other LASSO based method. Model performance was assessed in a simulation study with random data generated via symmetric local logistic regression models and Gibbs sampling. Main outcome variables were the Structural Hamming Distance and the Youden Index. We applied the results of the simulation study to a real-life data with functioning data of patients having head and neck cancer.

Results: Bootstrap aggregating as incorporated in the Bolasso algorithm greatly improved the performance in higher sample sizes. The number of bootstraps did have minimal impact on performance. Bolasso performed reasonable well with a cutpoint of 0.90 and a small penalty term. Optimal prediction for Bolasso leads to very conservative models in comparison with AIC, BIC or cross-validated optimal penalty terms.

Conclusion: Bootstrap aggregating may improve variable selection if the underlying selection process is not too unstable due to small sample size and if one is mainly interested in reducing the false discovery rate. We propose using the Bolasso for graphical modeling in large sample sizes.

3.1 Background

A common problem in contemporary biomedical research is the occurrence of a large number of variables that accompany relatively few observations. Thus, studying associations in high-dimensional data is not straightforward. Including all variables would result in a highly overparameterized model, computational complexity and unstable estimation of the associations [1]. This methodological problem has been solved for the domain of genomic medicine by using graphical modeling. Graphical models were identified as a promising new approach to modeling clinical data [2], and thereby the systems approach to health and disease.

A promising approach to describe such complex relationships is graphical modeling. Graphical models [3] provide a probabilistic tool to display, analyze and visualize the net-like dependence structures by drawing a graph describing the conditional dependencies between the variables. A graphical model consists of nodes representing the variables and edges representing conditional dependencies between the variables. In order to understand graphical models it is important to understand the concept of conditional independence. Two variables X and Y are considered conditional independent given Z, if $f(x|y,z) = f(x|z)$. Thus, learning information about Y does not give any additional information about X, once we know Z.

Beyond association, this method has also been developed for estimating causal effects [4]. Recently, graphical modeling has been outlined as a tool for investigating complex phenotype data, specifically for the visualization of complex associations [5], dimension reduction, comparison of substructures and the estimation of causal effects from observational data [6]. Until now, the main focus of research was on building Gaussian graphical models for continuous multivariate data following a multivariate normal distribution [7]. A popular way to built Gaussian graphical models are covariance selection methods [8]. These methods are used to sort out conditionally independent variables. They aim to identify the non-zero elements in the inverse of the covariance matrix since the non-zero entries in the inverse covariance matrix correspond to conditional dependent variables. However, this method is not reliable for high-dimensional data, but can be improved by concentrating on low-order graphs [9].

Another approach to identifying the non-zero elements in the inverse covariance matrix has been proposed by Meinshausen and Buehlmann [10]. They propose the Least Absolute Shrinkage and Selection Operator (LASSO) [11] as a variable selection method to identify the neighborhood of each variable, thus the non-zero elements. A neighborhood is the set of predictor variables corresponding to non-zero coefficients in a prediction model by estimating the conditional independence separately for each variable. Meinshausen and Buehlmann showed that this method is superior to common covariance selection methods, in particular if the number of variables exceeds the number of observations. They also proved that the method asymptotically recovers the true graph.

The LASSO was originally proposed for linear regression models and has become a popular model selection and shrinkage estimation method. LASSO is based on a penalty on the sum of absolute values of the coefficients (l_1-type penalty) and can be easily adapted to other settings, for example Cox regression [12], logistic regression [13, 14, 15] or multinomial logistic regression [16] by replacing the residual sum of squares by the corresponding negative log-likelihood function. Important progress has been made in recent years in developing computational efficient and consistent algorithms for the LASSO with good properties even in high-dimensional settings [17, 18]. The so-called graphical lasso by Friedman et al. [19] uses a coordinate descent algorithm for the LASSO regression to estimate sparse graphs via the inverse covariance matrix. They describe the connection between the exact problem and the approximation suggested by Meinshausen and Bühlmann [10]. However, the determination of the right amount of penalization for these methods has remained a main problem for which no satisfactory solution exists [20].

The current methodology primarily provides solutions for continuous data. The relationship of binary data is difficult to identify using classical methods. Building binary graphical models for dichotomous data is based on the corresponding contingency tables and log-linear models [21]. The interaction terms are used to control for conditional dependencies. With a growing number of variables model selection becomes computationally demanding and quickly exceeds feasibility, thereby making the method difficult to adapt to high-dimensional data. For a fully saturated log-linear model one would need $2 \cdot p$ parameters, with p being the number of variables. A common solution is to reduce the problem to first-order interaction where conditional independence is determined by first-order interaction terms.

The properties of the LASSO for logistic regression have recently been investigated. Van de Geer [22] focused on the prediction error of the estimator and not on variable selection. She proposed a truncation of the estimated coefficients to derive consistent variable selection. Bunea [23] showed the asymptotic consistency of variable selection under certain conditions for l_1-type penalization schemes.

The adaptation of local penalized logistic regression to graphical modeling has been proposed by Wainwright [24]. Under certain conditions on the number of variables n , the number of nodes p and the maximum neighborhood size, the l_1-penalized logistic regression for high-dimensional binary graphical model selection gives consistent neighborhood selection [24, 25]. Wainwright et al. showed that a logarithmic growth in n relative to p is sufficient to achieve neighborhood consistency. Another new approach is based on an approximate sparse maximum likelihood (ASML) problem for estimating the parameters in a multivariate binary distribution. Based on this approximation a consistent neighborhood could be selected and a sensible penalty term can be identified [17].

However, when analyzing high-dimensional categorical data the main problem that there is no rationale for the choice of the amount of penalization controlled by the value of the penalty term for consistent variable selection still remains [20]. A possible solution might be to adapt bootstrap aggregating to these problems. Bootstrap aggregating (bagging) generates multiple versions of a classifier and aggregates the results to get a single enhanced classifier. By making bootstrap replicates of the original data multiple versions are formed, each acting as single learning data for a classification problem. Also, for linear regression it has been shown that bagging provides substantial gains in accuracy for variable selection and classification [26]. This idea has been carried further by Bach [27], resulting in the Bolasso (bootstrap-enhanced least absolute shrinkage operator) algorithm for variable selection in linear regression. Here, the LASSO is applied to several bootstrapped replications of a given sample. The intersection of each of these models leads to consistent model selection.

In this paper we adapted the method of Meinshausen and Bühlmann to binary data and used the LASSO for logistic regression to identify the conditional dependence structure. We applied bagging to improve variable selection, hence adapted the Bolasso. Performances are tested on a data set with known structure. This data set was simulated by Gibbs sampling [28]. We also applied graphical modeling methods to real-life data.

Objective of this paper was to examine the performance of the Bolasso to the development of graphical models for high dimensional binary data with various values for the penalty term and various numbers of bootstraps. We hypothesized that the performance of Bolasso is superior to competing LASSO based methods to identify graphical models. Specifically, the hypothesis was that the choice of the penalty is not critical as long as it is chosen sensibly, i.e. corresponding to a reasonable number of selected variables.

3.2 Methods

3.2.1 Data generation

This section presents an approach to simulate high-dimensional binary data from a given distribution and dimension by analyzing the results on a data set with known dependence structure. This analysis is performed in order to investigate the performance of the proposed methods. All calculations are done using the statistical computing software R (V 2.9.0) [29]. We propose to generate the data via symmetric local logistic regression models and Gibbs sampling [28] as follows:

Algorithm 1 (Data generation)

1. *Define the* $p \times p$ *matrix* M *of odds ratios as:*
   $diag(M) = p_{ii}, i = 1, \ldots, p$ *with* p_{ii} *the baseline odds of variable* $X^{(i)}$
   $m_{ij} = p_{ij}$ *with* p_{ij} *as the corresponding odds ratio of* $X^{(i)}$ *on* $X^{(j)}$ *and vice versa.*

2. *Start with* $k = 0$.

3. *Choose starting values* $x^k = x_1, \ldots, x_p$ *according to* $diag(M)$.

4. *For each* i *in* $1, \ldots, p$ *generate new* x_i^{k+1} *from a Bernoulli distribution* $B(p_i^*)$ *according to* $logit(p_i^*) = \sum_{j \neq i} m_i^j \cdot x_j^{(k)}$.

5. *Repeat for* $k = k + 1$.

After a burn-in phase the $x_i^{(k+1)}$ will reflect the true underlying binary distribution generating $X = (X^{(1)}, \ldots, X^{(p)}) \in 0, 1^p$. We chose a burn-in phase of 5000 iterations.

3.2.2 Real-Life Data: Aspects of functioning in Head and Neck cancer patient

We evaluated the method to data measuring aspects of functioning of patients having head and neck cancer (HNC). The data originated from a cross-sectional study with a convenience sample of 145 patients with HNC. The data has previously been used for graphical modelling and has already been accepted for publication [30].

The patients had at least one cancer at one of the following locations: oral region, salivary glands, oropharynx, hypopharynx or larynx. Human functioning for each of the patients were assessed using the International Classification of Functioning, Disability and Health (ICF) as endorsed by the World Health Organization in 2001 [31]. The ICF provides a useful framework for classifying the components of health and consequences of a disease and can be used. According to the ICF the consequences of a disease may concern body functions (b) and structures (s), the performance of activities and participation (d) in life situations depending on environmental factors (e). Thirty-four aspects of functioning were assessed for each of the patients 12 from the component Body Functions, three from the component Body Structure, 15 from the component Activity and Participation and another 4 categories from the component Environmental factors. For better interpretation of the graphs we show the 34 ICF categories together with a short explanation in Table 3.1.

Table 3.1 – *Short description of the ICF categories used for the graphical models on the HNC data.*

ICF Code	ICF Code description
b130	Energy and drive functions
b280	Sensation of pain
b310	Voice functions
b320	Articulation functions
b340	Alternative vocalization functions
b440	Respiration functions
b450	Additional respiratory functions
b460	Sensations associated with cardiovascular and respiratory functions
b510	Ingestion functions
b515	Digestive functions
b530	Weight maintenance function
b710	Mobility of joint functions
d175	Solving problems
d310	Communicating with - receiving - spoken messages
d315	Communicating with - receiving - nonverbal messages
d330	Speaking
d335	Producing nonverbal messages
d350	Conversation
d360	Using communication devices and techniques
d550	Eating
d560	Drinking
d570	Looking after one´s health
d720	Complex interpersonal interaction
d760	Family relationship
d770	Intimate relationship
d850	Remunerative employment
d920	Recreation and leisure
s320	Structure of mouth
s430	Structure of respiratory system
s710	Structure of head and neck region
e125	Products and technology for communication
e225	Climate
e310	Immediate family
e580	Health services, systems and policies

3.2.3 Principles of graphical models

Consider $X = (X^{(1)}, \ldots, X^{(p)}) \in 0, 1^p$ as a p-dimensional vector of binary random variables. One way to represent the association structure between the elements of in a random sample of i.i.d. replicates is an undirected binary graph. A graph $G(\upsilon, \epsilon)$ consists of a finite set of nodes υ, representing the elements of X, and edges ϵ between these nodes. Each edge stands for an existing conditional dependence between two nodes. Hence, graphical modeling is based on the concept of conditional dependence and conditional independence. To understand graphical models it is fundamental to understand both of these concepts. Two events X and Y are independent, if $P(X \cap Y) = P(X) \cdot P(Y)$. Two events X and Y are conditional independent given Z if $P(X \cap Y|Z) = P(X|Z) \cdot P(Y|Z) \Leftrightarrow X \perp Y|Z$. The relationship $X \perp Y|Z$ is represented in figure 3.1.

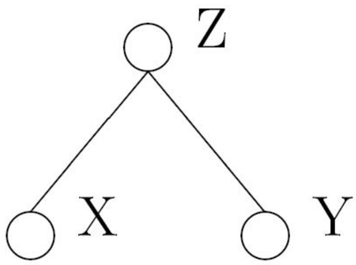

Figure 3.1 – *An example for a simple graphical model*

A concept to describe a graphical model is via the neighborhood of each node. The neighborhood of node $X^{(a)}$ ne_a is defined as the smallest subset of υ, so that $X^{(a)}$ is conditionally independent of all remaining variables, thus the neighborhood ne_a is defined as:

$$X^{(a)} \perp \left\{ X^{(i)}; \forall X^{(i)} \in \upsilon/ne_a \right\} | ne_a \tag{3.1}$$

Two approaches define the edge set ϵ. First, an edge between x and y exists if and only if both nodes are an element in the opposite neighborhood, i.e. following the AND-rule:

$$\epsilon = \{(x,y)|x \in ne_y \wedge y \in ne_x\} \tag{3.2}$$

A less conservative, asymmetrical estimate of the edge set of a graph is given by the OR-rule:

$$\epsilon = \{(x,y)|x \in ne_y \vee y \in ne_x\} \tag{3.3}$$

A different definition of a neighborhood allows for a practical approach. For each node in υ consider optimal prediction of $X^{(a)}$ given all remaining variables. Let $\beta^a \in \Re^{(p-1)}$ be the vector of coefficients for optimal prediction of $X^{(a)}$. The set of non-zero coefficient is identical to the set of neighbors of $X^{(a)}$, thus:

$$ne_a = b \in \upsilon : \beta_b^a \neq 0 \tag{3.4}$$

We can regard this as a subset selection problem in a regression setting and to the detection of zero coefficients. The use of shrinkage as the subset selection tool used to identify the neighborhood and to estimate β_b^a has become popular in recent years.

3.2.4 Principles of LASSO for logistic regression

Given a set of p explanatory $X^{(1)}, \ldots, X^{(p)}$ and a binary outcome Y the goal of logistic regression is to model the probability $\pi_i = p(y_i = 1|x_i)$ by

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = x_i'\beta \Leftrightarrow \pi_i \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} \tag{3.5}$$

The maximum likelihood estimates of β can be found by setting the first derivative of log-likelihood function equal to zero, thus

$$\frac{\partial l(\beta)}{\partial \beta} = s(\beta) = \sum_{i=1}^{N} x_i \left(y_i - \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)} \right) = 0 \tag{3.6}$$

The LASSO is a penalized regression technique defined as a shrinkage estimator [11]. It adds a penalty term to equation 3.6, thus 3.6 is to be minimized subject to the sum of the absolute coefficients being less than a certain threshold value. Using the absolute values as condition yields shrinkage in some coefficients and simultaneously may set other coefficients to zero as has been shown by Tibshirani [11]. For each choice of the penalty parameter a stationary solution exists often visualized as a regularization path, i.e. the penalized coefficients over all penalty terms. LASSO reduces the variation in estimating β. Formally, the penalized logistic regression problem is to minimize:

$$\sum_{i=1}^{N} x_i \left(y_i - \frac{\exp(x_i'\beta^{LASSO})}{1 + \exp(x_i'\beta^{LASSO})} \right) + \lambda \sum_j |\beta_{(j)}^{LASSO}| \tag{3.7}$$

A mathematically equivalent expression of this problem is the formulation as a constrained regression problem, that is minimizing

$$\sum_{i=1}^{N} x_i \left(y_i - \frac{\exp(x_i'\beta^{LASSO})}{1 + \exp(x_i'\beta^{LASSO})} \right). \tag{3.8}$$

subject to $\sum_j |\beta_{(j)}^{LASSO}| < t$.

In this paper, the variables are first standardized to zero mean and variance one. The penalty term t can be used to control the number of predictors, i.e. the size of the neighborhood.

3.2.5 Binary graphical model using single LASSO regressions

The first method considered here to construct graphical models is based on an optimal penalty term to identify an optimal neighborhood. We consider three procedures for selecting optimal penalty, namely cross-validation (CV), Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). These approaches have although been suggested in recent publications by Viallon [32] and Wang [33]. The first one is proposed by Goeman in the context of Cox regression [34], while the latter two are computational more efficient and well-known classical tools. BIC has shown to be superior as suggested by Yuan and Lin and demonstrated superior performance through simulation studies in Gaussian graphical modes [35, 36]. Other possible approaches include the methods by Wainwright [24] and Banerjee [17] which control the rate of falsely detected edges in the graph. However, these penalties are too conservative as outlined in Ambroise et al. [37]. The LASSO and cross-validation are calculated using the efficient gradient-ascent algorithm as proposed by Goeman and are implemented in the 'penalized' package [38].

For all possible penalty terms, the performance of the resulting models is assessed either by cross-validation, AIC or BIC. The algorithm to identify a binary graphical model then proceeds as follows:

Algorithm 2 (Graphical Model)

1. *Estimate the coefficients* $\hat{\beta}^{LASSO}$ *in local penalized logistic regression models using each variable as outcome and the remainder as predictors for each* $X^{(i)}$ *corresponding to an optimal penalty term* t*.*

2. *For each variable* a *define the neighborhood* ne_a^* *as the set of variables* b *corresponding to non-zero penalized coefficients*

$$\hat{\beta}_b^a : ne_a^* = \{b \in \upsilon : \hat{\beta}_b^a \neq 0\}.$$

3. *Define the set of conditional relationships (the edge set)* E *as:*

$$E = \{(a,b)|a \in ne_b^* \vee b \in ne_a^*\}.$$

3.2.6 Binary graphical model using Bolasso

Another method to construct binary graphical models is based on the Bolasso algorithm which takes advantage of bootstrap aggregating. Bootstrap aggregating, also called 'bagging', generates multiple versions of a predictor, e.g. a coefficient in a generalized linear model, or classifier. It constitutes a simple and general approach to improve an unstable estimator $\theta(X)$ with X being a given data set. Bagging is based on the concept of resampling the data $\{(y_n, x_n), n = 1, \ldots, p\}$ by drawing new pairs $\{(y_n^*, x_n^*)\}$ with replacement from the p original data pairs. For simple classification the bootstrap aggregating algorithm proceeds as follows:

Algorithm 3 (Bootstrap Aggregating)

1. *Generate a bootstrap sample* b^* *with replacement from the original data. Repeat this process* B *times.*

2. *For each data sample* b^* *calculate the classifier* $\theta^*(b^*)$*.*

3. *Count the times an object is classified* $\mu_x = \sum\limits_{i=1,\ldots,B} \theta_i^*(b_i^*)$*.*

4. *Define the set of classified objects as* $S = \{y : \mu_y \geq \pi_{cut}\}$ *with* $0 \leq \pi_{cut} \leq 1$*.*

Using Bolasso as a basis, we can now construct graphical models. The algorithm proceeds as follows:

Algorithm 4 (Graphical Models using Bolasso)

1. *Generate a bootstrap sample b^* with replacement from the original data. Repeat this process B times.*

2. *For each b^*, estimate the coefficients $\hat{\beta}^{LASSO}$ in local penalized logistic regressions using each variable as outcome variable and the remainder as predictors for a penalty term t.*

3. *For each variable a define the neighborhood ne_a^* as the set of variables b corresponding to non-zero penalized coefficients*

$$\hat{\beta}_b^a : ne_a^* = \{b \in \upsilon : \hat{\beta}_b^a \neq 0\}.$$

4. *Calculate the percentage variable b is in the neighborhood of a in each bootstrap sample b^* as μ_b^a.*

5. *Define the neighborhood of variable a as: $ne_a^* = \{b|\mu_b^a \geq \pi_{cut}\}$.*

6. *Define the set of conditional relationships (the edge set) E as:*

$$E = \{(a, b)|a \in ne_b * \vee b \in ne_a^*\} .$$

Three parameters have to be chosen here:

1. B: number of bootstraps
2. t: penalty parameter
3. π_{cut}: cut-off value for the definition of neighborhood

In our study, we investigated the influence of these parameters for different sample sizes in a study based on simulated data as described earlier. We also considered using a cross-validated penalty as a basis for the Bolasso and refer to this approach by Bolasso-CV.

3.2.7 Assessment of performance

We analyzed the performance of the methods by comparing the identified structure with a predefined known structure. Thus, each edge yielded by one method can be either correct or incorrect. A falsely identified (false positive) edge is an edge which is identified by one or both of the two methods but does not exist in the predefined structure. A falsely not identified (false negative) edge is an edge which is not identified by one or both of the two methods but does exist in the predefined structure. A correctly identified (true positive) edge is an edge which is identified by one or both of the two methods and which exists in the predefined structure. Likewise, a true negative edge is an edge correctly identified as missing.

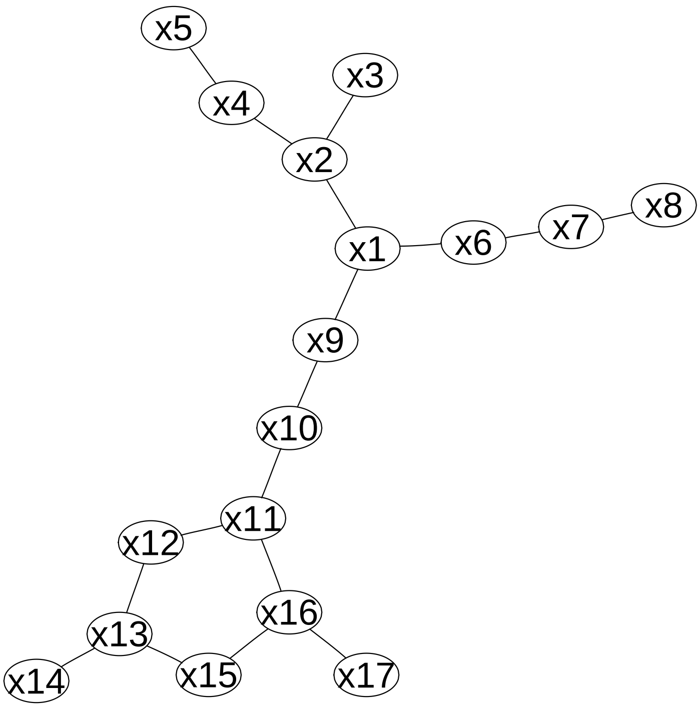

Figure 3.2 – *The simulation setting for assessing the performance of the methods*

We report the Structural Hamming Distance (SHD) and the Youden index (J). The SHD between two graphs is the number of edge insertions or deletions needed to transform one graph into the other. Thus, the number of changes needed to transform the graphical model identified by one or both of the two methods to the known structure defined by the matrix M. The SHD measures the performance of LASSO and Bolasso by counting the number of false positive and false negative edges. It may occur that bagging causes the exclusion of all edges yielding a SHD equal to the number of true edges. This might reduce the error rate, but an empty model without edges is not always desirable, even if it has a low error rate. In order to assess both, the ability to find true positive and true negative edges, the Youden Index is more appropriate.

The Youden index is a function of the sensitivity (q) and the specificity (p) of a classifier and is defined as:

$$J = q + p - 1. \tag{3.9}$$

Sensitivity is the proportion of true positive edges among all identified edges and specificity the proportion of true negative edges among all not identified edges. Smaller values of the SHD indicate better choice, as do larger values of the J. Thus, a choice is to be preferred that yields small SHD at large J values.

We investigated the performances in a simulation setting which was motivated by a graphical model for real-life data [5] (see figure 3.2). In this study functioning data for patients in the post-acute setting were analyzed using graphical modeling. The setting mimics a found subgraph in this graphical model. We additionally added two random variables having no interaction with the remainders to imitate a realistic scenario. The model in figure 3.2 corresponds to a particular matrix of odds ratios, e.g. a smaller model with only 6 variables can be expressed by the matrix M:

$$M = \begin{pmatrix} 1 & 2 & 2 & 1 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

We chose the penalty term to be either cross-validated, AIC or BIC optimal or to correspond to a certain neighborhood size ranging from one to the maximum neighborhood size, i.e. $l \in (1, 2, \ldots 18)$ in the original data. In addition, we varied the number of bootstrap replicates $B \in (40, 80, 120, 160, 200)$, the threshold for a variable to be included in a neighborhood $\pi_{cut} \in (0.90, 0.95, 0.99, 1.00)$ and the sample size $n \in (50, 100, 200, 500, 1000)$. Usually, B lies in the range of 50. However, the best choice is not clear, e.g. Bach investigated also $B = 256$, and may be important as the method itself is unstable. The choice of thresholds was also motivated by the work of Bach who proposed the soft Bolasso with a threshold of 0.90 as opposed to the Bolasso with a threshold of 1.00. We chose a wide range of values for B, thresholds and sample size to simultaneously study the performance of B and thresholds in small samples and in big samples. Ideally, for a sample size of 1000 the methods should perform with negligible error. In order to estimate model performances dependent on the parameters π_{cut}, B and l and the interaction between π_{cut} and l we calculated generalized linear models with either SHD or J as outcome variable.

3.3 Result

3.3.1 Simulation Results

We calculated the Structural Hamming Distance and Youden Index for the simulation setting for each combination of B, π_{cut} and l. We give the detailed results in table form in the electronic supplement (see additional file 1 and additional file 2: Summary statistics for the simulation study). Using generalized linear model yielded the following optimal regularization for minimal SHD: $\pi_{cut} = 0.90$, $B = 200$ and $l = 5$. For a maximal J a larger neighborhood size is preferred: $\pi_{cut} = 0.90$, $B = 200$ and $l = 10$. It turned out, that the number of bootstrap replicates B has the least influence on model performance with higher B performing slightly better.

We give a summary of the results for SHD in figure 3.3 and for J in figure 3.4 for varying sample size. The figures show boxplots for all methods, namely the Lasso-CV, the Lasso-AIC, the Lasso-BIC, the Bolasso-CV and the Bolasso with optimal

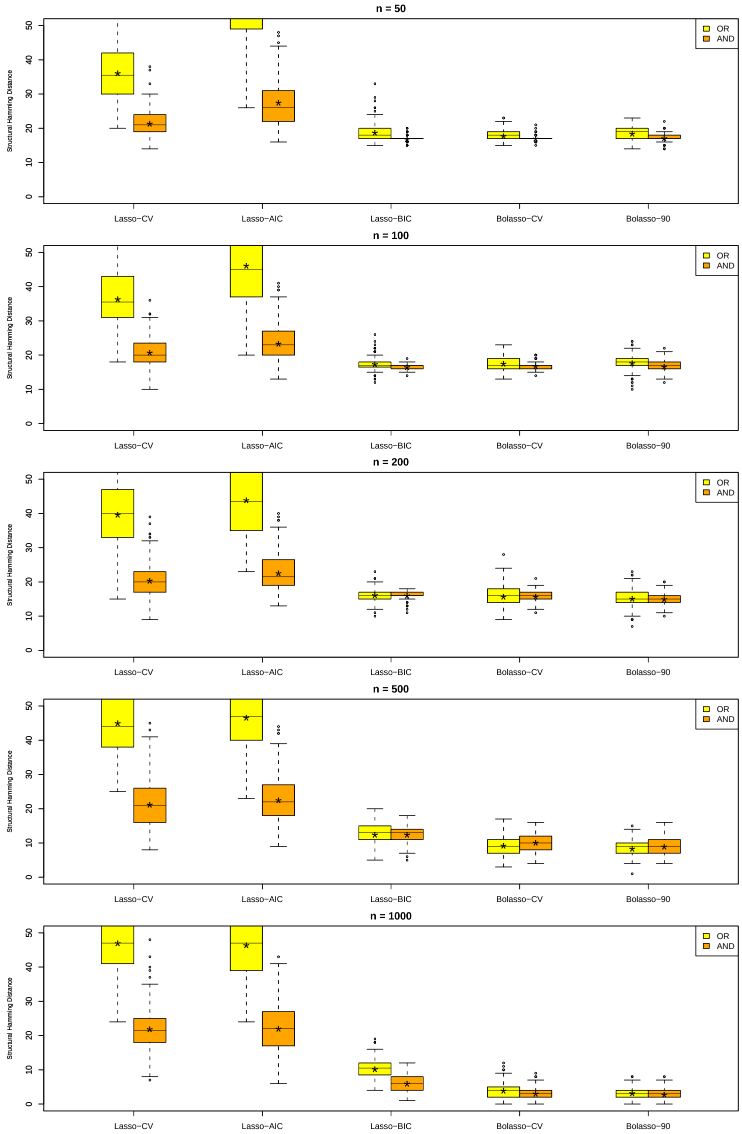

Figure 3.3 – *Boxplot of Structural Hamming Distance (SHD) for all investigated approaches, different definitions of neighborhood (OR-rule and AND-rule) and sample sizes 50, 100, 200, 500 and 1000. Yellow marks the OR-rule and orange the AND-rule. Lasso-CV represents cross-validated optimal penalty, Lasso-AIC represents AIC optimal penalty, Lasso-BIC represents BIC optimal penalty, Bolasso-CV represents Bolasso with cross-validated optimal penalty and Bolasso-90 represents a Bolasso with a cut of 90% and neighborhood size 5.*

neighborhood size and $\pi_{cut} = 0.90$, the Bolasso-90. For each method we applied two different neighborhood definitions corresponding to the AND-rule and to the OR-rule. In the boxplots * marks the mean performance.

Considering SHD as the main outcome renders Lasso-CV and Lasso-AIC as clearly inferior. For smaller sample sizes Lasso-BIC, Bolasso-CV and Bolasso-90 reject almost or all edges Leading to a null model with SHD equal to 19. For sample sizes greater than 500 Bolasso-CV and Bolasso-90 are clearly superior to Lasso-BIC. Additionally, for each approach the AND-rule was superior to the OR-rule for most

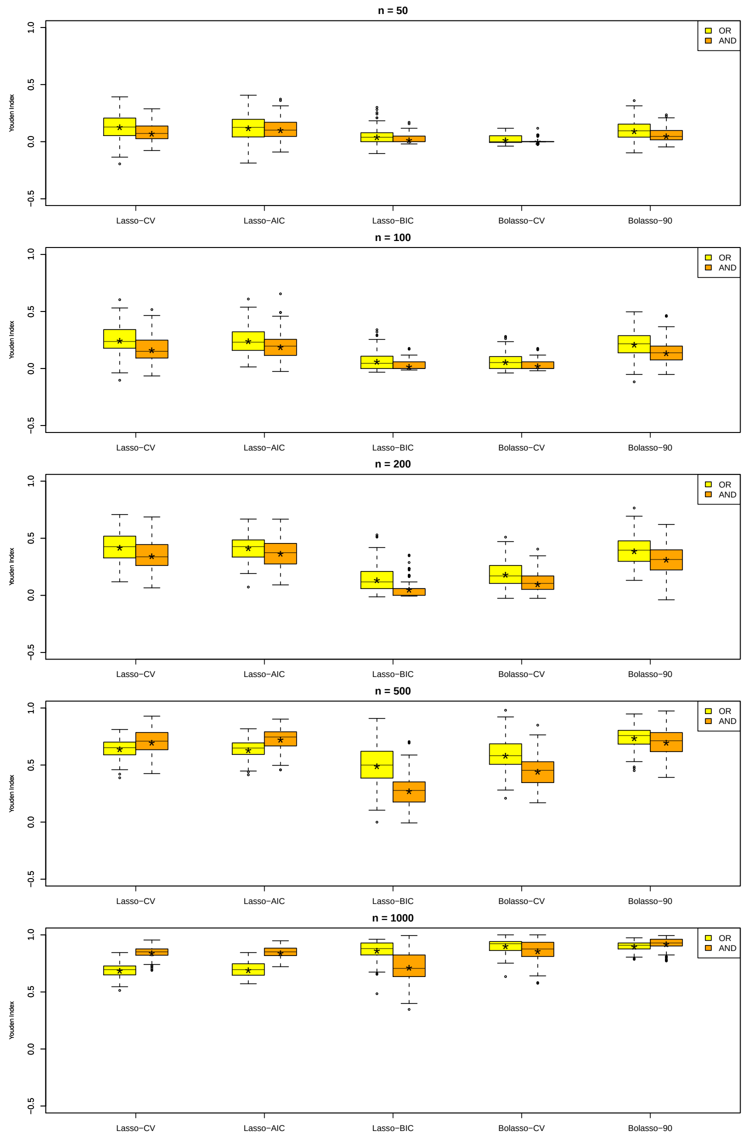

Figure 3.4 – *Boxplot of Youden Index (J) for all investigated approaches, different definitions of neighborhood (OR-rule and AND-rule) and sample sizes 50, 100, 200, 500 and 1000. Yellow marks the OR-rule and orange the AND-rule. Lasso-CV represents cross-validated optimal penalty, Lasso-AIC represents AIC optimal penalty, Lasso-BIC represents BIC optimal penalty, Bolasso-CV represents Bolasso with cross-validated optimal penalty and Bolasso-90 represents a Bolasso with a cut of 90% and neighborhood size 10.*

sample sizes.

A similar result can be seen when considering the Youden Index with the exception that Lasso-CV and Lasso-AIC are real contenders here, as they are not as conservative as the others, such gaining performance regarding sensitivity.

3.3.2 HNC Data

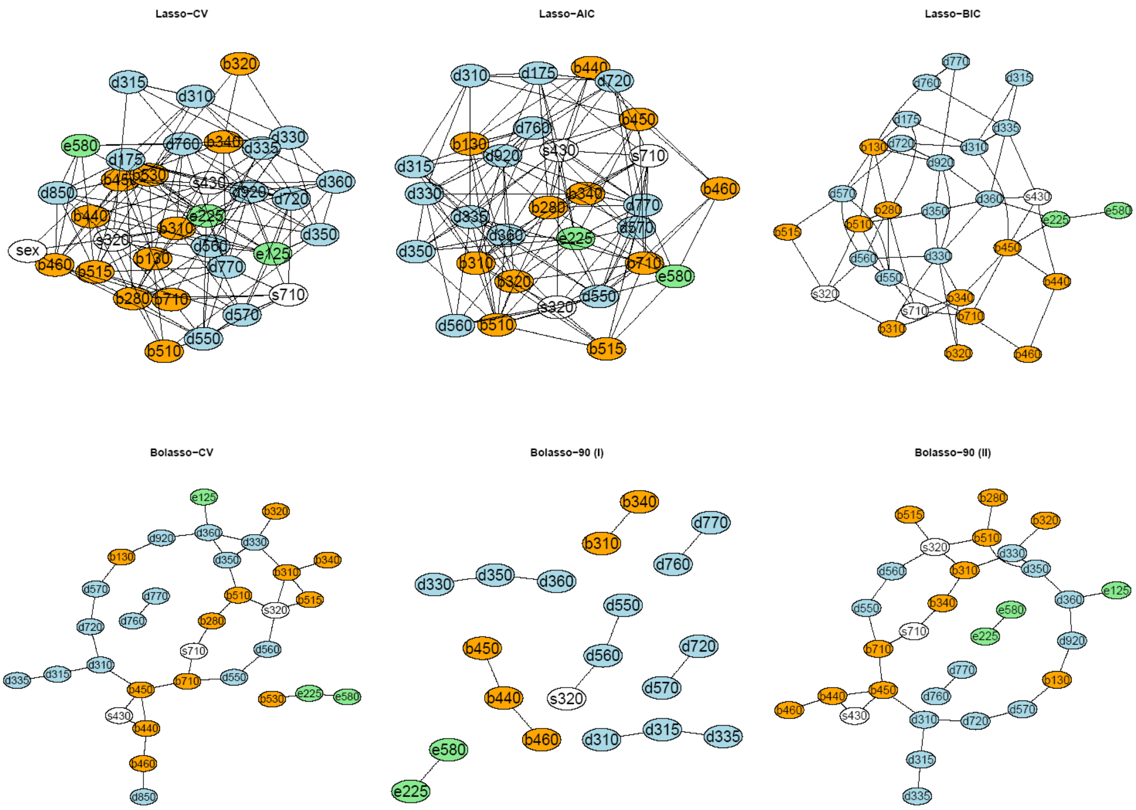

Figure 3.5 – *Graphical models for the real-life data using the AND-rule. ICF categories from the component Body function (orange), Body structure (white), Activities and participation (blue) and Environmental factors (green). Please, find the full descriptions of the ICF categories in Table 3.1.*

Using these results we applied the method to the HNC data set using both the AND- and OR-rule. We give the results for the CV, the AIC and the BIC optimal penalty and for the Bolasso with a cut of 90%, 200 bootstraps and a neighborhood size of 5 (for optimal SHD), resp. 10 (for optimal J) in figure 3.5 and 3.6. The color of the nodes correspond to the different ICF components: ICF categories from the component Body function are orange, Body structure white, Activities and participation blue and Environmental factors green. The full descriptions of the ICF categories are in Table 3.1.

In all models similar aspects can be seen. The CV and AIC optimal penalty term leads to a very complicated model, while the BIC criteria yielded reasonable results in terms of interpretability. The Bolasso-90 is the most conservative approach while using Bolasso-CV yielded similar results than the Lasso-BIC.

As a case in point, we describe the model for the Bolasso-90 with a neighborhood size of 10, i.e. the model with the highest performance regarding the Youden Index. Similar to Becker et al. [30] we identified a circle-like association around the speaking capability, i.e. between d330 *Speaking*, b310 *Voice functions*, b320 *Articulation func-*

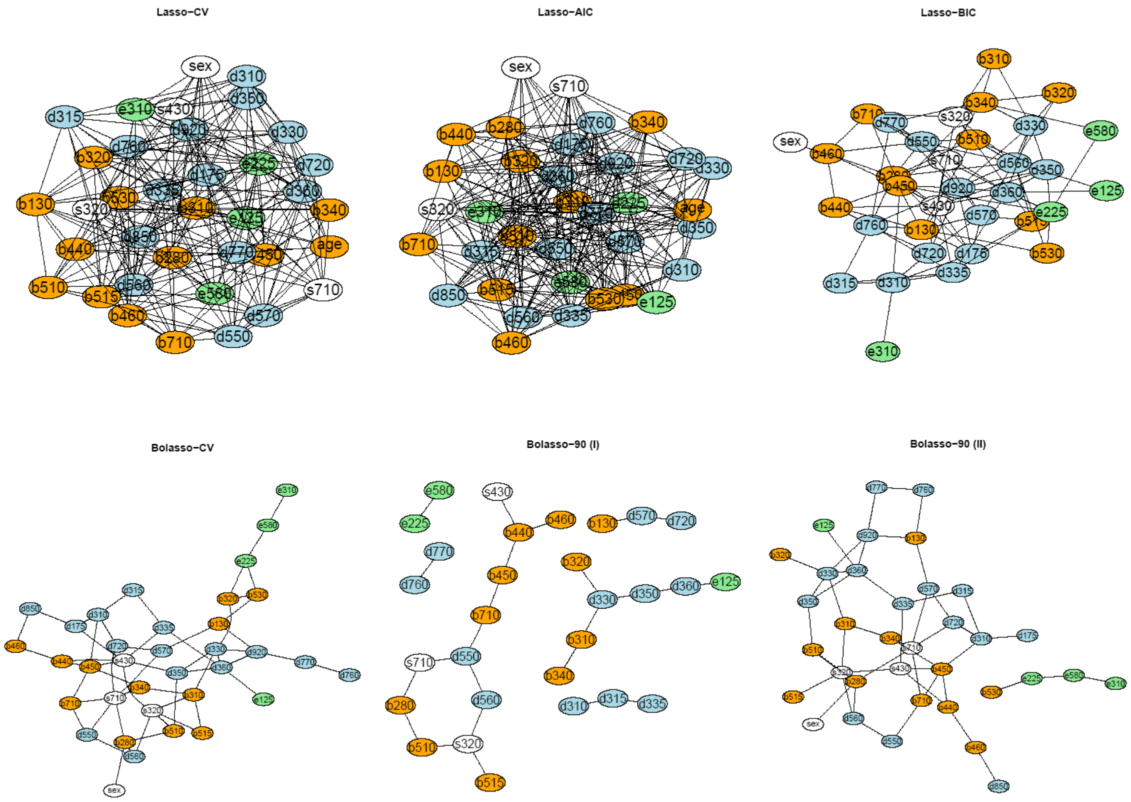

Figure 3.6 – *Graphical models for the real-life data using the OR-rule. ICF categories from the component Body function (orange), Body structure (white), Activities and participation (blue) and Environmental factors (green). Please, find the full descriptions of the ICF categories in Table 3.1.*

tions, s320 *Structure of mouth*, b510 *Ingestion functions*, d350 *Conversation*, d360 *Using communication devices and techniques*. The latter had further associations to e125 *Products and technologies for communication* and d920 *Recreation and leisure*. The category s320 *Structure of mouth* had a meaningful connection to d560 *Drinking* which was further connected to d550 *Eating*. Furthermore, b510 *Ingestion functions* had an association to b280 *Sensation of pain*. On the left side of the graph we have a group around respiration functions, namely b440 *Respiration functions*, b450 *Additional respiratory functions*, b460 *Sensations associated with cardiovascular and respiratory functions* and s430 *Structure of respiratory system*. A further like path could be visualized between the categories d335 *Producing nonverbal messages*, d315 *Communicating with - receiving - nonverbal messages*, d310 *Communicating with - receiving - spoken messages*, d720 *Complex interpersonal interaction*, d570 *Looking after one´s health* and b130 *Energy and drive functions*. The big circle is closed by the connection of b130 and d920 *Recreation and leisure*.

Many of these association structures were also present in the original work and are discussed in detail there [30].

3.4 Discussion

We compared the performance of the Bolasso to the development of graphical models for high-dimensional data with known dependency structure. One of the main points of critique for graphical models is that the retrieved structures might not be statistically stable, since the results might depend on the choice of model parameters, and are susceptible to small changes in the data set [6]. Interestingly, we found that using a BIC penalty and Bolasso were both able to correctly identify predefined existing dependency structures.

We have analyzed several LASSO based methods to derive graphical models in the presence of binary data and compared their performance in detecting known dependency structures. All methods are taking advantage of penalized logistic regression models as a tool to identify the explicit neighborhoods. We could show that bootstrap aggregating can substantially improve the performance of model selection, especially in the case of large samples. Arguably, LASSO is inferior in certain situations because in a given LASSO coefficient path the optimal solution might not exist. A LASSO coefficient path is given by the coefficient value of each variable over all penalty terms. In contrast, bagging opens the window for a whole class of new models, because it selects all variables intersecting over the bootstrap samples. The intersection itself must not be a solution in any of the bootstrap samples.

In LASSO, the choice of the penalty often determines the performance of the model. Thus, the correct choice of the penalty term is important. However, in our study, the initial choice of penalization had surprisingly little impact on the performance of the model if bagging was applied and the penalty was chosen in sensibly. Similar results have been obtained with stability selection [20]. Stability selection is also based on bootstrap in combination with (high-dimensional) selection algorithms. It applies resampling to the whole LASSO coefficient path and calculates the probability for each variable to be selected when randomly resampling from the data.

In our study, bagging largely improved the performance of LASSO, but only by reducing the number of false positive and false negative edges. This, however, might lead to a conservative and underspecified model with a low number of edges, if any, especially in small samples.

Although the choice of the penalty term is not crucial when bagging is applied, a cut-off value for the definition of neighborhood has to be defined, and this arbitrary choice will determine the size of the graphical model. Reducing the cut-off value would include more variables at the expense of a higher false positive rate.

This study illustrates that in graphical modeling, it is essential not only to control the number of false positive and false negative edges but also the ability of a method to identify true positive edges.

3.5 Conclusion

Bootstrap aggregating improves variable selection if the underlying selection process is not too unstable, e.g. due to small sample sizes. These properties have been shown on simulated data using various parameters. As a consequence, we propose using Bolasso for graphical modeling in large sample size cases as a real contender to the classical neighborhood estimation methods.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RS computed the study; EG and UM contributed to the design of the study; all authors contributed to the analysis and writing. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Emmi Beck for manuscript revision and Sven Becker for providing the data for the real-data study.

This study was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) (GR 3608/1-1).

Literaturverzeichnis

[1] Leo Breiman. Heuristics of Instability and Stabilization in Model Selection. *The Annals of Statistics*, 24(6):2350–2383, 1996.

[2] CL Tsai and CA Camargo. Methodological considerations, such as directed acyclic graphs, for studying äcute on chronicd̈isease epidemiology: Chronic obstructive pulmonary disease example. *Clinical Epidemiology*, 62:982–990, 2009.

[3] David Edwards. *Introduction to Graphical Modelling.* Springer, New York, 2000.

[4] M. A. Hernan and J. M. Robins. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*, 17(4):360–72, 2006.

[5] R Strobl, G Stucki, E Grill, M Müller, and U Mansmann. Graphical models illustrated complex associations between variables describing human functioning. *J Clin Epidemiol*, (62):922–33, 2009.

[6] M. Kalisch, B. A. Fellinghauer, E. Grill, M. H. Maathuis, U. Mansmann, P. Buhlmann, and G. Stucki. Understanding human functioning using graphical models. *BMC Med Res Methodol*, 10:14, 2010.

[7] Frederick Wong, Christopher K. Carter, and Robert Kohn. Efficient estimation of covariance selection models. *Biometrika*, 90(4):809–830, 2003.

[8] A.P. Dempster. Covariance Selection. *Biometrics*, 28(1):157–175, 1972.

[9] Anja Wille and Peter Bühlmann. Low-order conditional independence graphs for inferring genetic networks. *Statistical Applications in Genetics and Molecular Biology*, 5(1):1–32, 2006.

[10] Nicolai Meinshausen and Peter Buehlmann. High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*, 34:1436–1462, 2006.

[11] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58:267–288, 1996.

[12] R. Tibshirani. The lasso method for variable selection in the Cox model. *Stat Med*, 16(4):385–95, 1997.

[13] Alexander Genkin, David D. Lewis, and David Madigan. Large-Scale Bayesian Logistic Regression for Text Categorization. *Technometrics*, 49(3):291–304, 2007.

[14] Justin Lokhorst. The LASSO and Generalised Linear Models. *Honours Projects*, Department of Statistics. South Australia, Australia: The University of Adelaide; 1999.

[15] S. K. Shevade and S. S. Keerthi. A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics*, 19(17):2246–53, 2003.

[16] B. Krishnapuram, L. Carin, M. A. Figueiredo, and A. J. Hartemink. Sparse multinomial logistic regression: fast algorithms and generalization bounds. *IEEE Trans Pattern Anal Mach Intell*, 27(6):957–68, 2005.

[17] Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data. *Journal of Machine Learning Research*, 9:485–516, 2008.

[18] Adam J Rothman, Peter J Bickel, Elizaveta Levina, and Ji Zhu. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2:494–515, 2008.

[19] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–41, 2008.

[20] Nicolai Meinshausen and Peter Buehlmann. Stability Selection. *Journal of the Royal Statistical Society: Series B*, 72:417–473, 2010.

[21] Alan Agresti. *Categorical Data Analysis*. Wiley, New Jersey, 2002.

[22] Sara van der Geer. High-dimensional generalized linear models and the Lasso. *The Annals of Statistics*, 36(2):614–645, 2008.

[23] Florentina Bunea. Honest variable selection in linear and logistic regression models via l1 and l1 + l2 penalization. *Electronic Journal of Statistics*, 2:1153–1194, 2008.

[24] Martin J. Wainwright, Pradeep Ravikumar, and John D. Lafferty. High-dimensional graphical model selection using l1-regularized logistic regression. *Advances in Neural Information Processing Systems*, 19:1465–1472, 2007.

[25] Pradeep Ravikumar, Martin J. Wainwright, and John D. Lafferty. High-dimensional Ising model selection using l1-regularized logistic regression. *The Annals of Statistics*, 38(3):1287–1319, 2010.

[26] L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.

[27] Francis R. Bach. Bolasso: Model consistent lasso estimation through the bootstrap. *preprint*, 2008.

[28] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 2 edition, 2004.

[29] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0.

[30] S. Becker, R. Strobl, A. Cieza, E. Grill, U. Harreus, and U. Tschiesner. Graphical modeling can be used to illustrate associations between variables describing functioning in head and neck cancer patients. *J Clin Epidemiol*, (64):885–92, 2011.

[31] WHO. International Classification of Functioning, Disability and Health: ICF, 2001.

[32] Vivian Viallon, Onureena Banerjee, Gregoire Rey, Eric Jougla, and Joell Coste. An empirical comparative study of approximate methods for binary graphical models; application to the search of associations among causes of death in French death certificates. *preprint*, http://arxiv.org/pdf/1004.2287, 2010.

[33] Pei Wang, Dennis L Chao, and Li Hsu. Learning networks from high dimensional binary data: An application to genomic instability data. *preprint*, http://arxiv.org/pdf/0908.3882, 2010.

[34] Jelle J. Goeman. L1 penalized estimation in the Cox proportional hazards model. *Biometrical Journal*, 52(1):70–84, 2009.

[35] Xin Gao, Daniel Q. Pu, Yuehua Wu, and Hong Xu. Tuning parameter selection for penalized likelihood estimation pf inverse covariance matrix. *preprint*, http://arxiv.org/pdf/0909.0934, 2010.

[36] Ming Yuan and Yi Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

[37] Christophe Ambroise, Julien Chiquet, and Catherine Matias. Inferring sparse Gaussian graphical models with latent structure. *Electronic Journal of Statistics*, 3:205–238, 2009.

[38] Jelle Goeman. penalized: L1 (lasso) and L2 (ridge) penalized estimation in GLMs and in the Cox model. *R package version 0.9-21*, 2008.

KAPITEL 4

DIMENSION REDUCTION IN HUMAN FUNCTIONING AND DISABILITY OUTCOMES RESEARCH: GRAPHICAL MODELS VERSUS PRINCIPAL COMPONENTS ANALYSIS

Jan Reinhardt[a,b,c,d,e], Bernd Fellinghauer[a,f], Ralf Strobl[d,e], Gerold Stucki[a,b]

[a] *Swiss Paraplegic Research (SPF), Nottwil, Switzerland*

[b] *Department of Health Sciences and Health Policy, University of Lucerne, and SPF, Nottwil, Switzerland*

[c] *Department of Sociology, University of Lucerne, Lucerne, Switzerland*

[d] *ICF Research Branch of WHO FIC CC Germany (DIMDI) at SPF Nottwil, Switzerland*

[e] *Institute for Health and Rehabilitation Sciences, Ludwig-Maximilians-Universität München, Munich, Germany*

[f] *Department of Statistics, Swiss Federal Institute of Technology, Zurich, Switzerland*

This Chapter was published as

Reinhardt JD, Fellinghauer BA, Strobl R, Stucki G: **Dimension reduction in human functioning and disability outcomes research: graphical models versus principal components analysis.** *Disabil Rehabil* 2010, **32**:1000-1010.

Abstract

Purpose: The complexity of rehabilitation outcomes in terms of functioning and disability leads to the need for dimension reduction in relation to specific research or clinical problems. Several statistical procedures are at hand. This article examines whether and to what extent principal component analysis (PCA) and graphical models differ in the dimension reduction of data based on the International Classification of Functioning, Disability and Health (ICF).

Methods: Using a data set of 1048 persons with spinal cord injury from 14 different countries as a case in point, this article compares the solutions in dimension reduction generated with a graphical model based on least average shrinkage selection operator (LASSO) regression on one hand and with a PCA on the other.

Results: Some factors extracted with the PCA properly match the clusters found with the graphical model, while in others less commonality can be found. In general, overlap ranges from 9 to 100% with 75% on average. Many of the discovered clusters or factors, i.e. dimensions, are compatible with the ICF structure, particularly in activity and participation.

Conclusion: Functioning is a multidimensional, complex and dynamic outcome. We recommend being careful with dimension reduction based on statistical procedures alone. Theoretical considerations and clinical significance should always guide statistics. Stability of emergent dimensions that are compatible with theoretical or clinical considerations is a most important point. At least two different statistical procedures for dimension reduction, for instance PCA and LASSO regression, should be applied to conservatively select those dimensions that stay stable with both procedures.

Keywords: *Human functioning, disability, ICF, dimension reduction, PCA, graphical models*

4.1 Introduction

While survival is the most crucial outcome of acute care, functioning or disability is the most important outcome following the rehabilitation of persons with health conditions. Human functioning of persons with health conditions [1, 2, 3] encompasses all body functions and structures, various activities and all forms of participation in social roles [4] and life tasks [5]. The concepts of health condition and diagnosis refer to clusters of symptoms and signs as proxies of underlying pathologies [6]. The concepts of functioning and disability refer to the lived human experience of people in relation to health conditions. Two persons with different health conditions can have similar functioning outcomes and two people with the same health condition can have completely different outcomes in functioning depending on personal and environmental factors [1]. The complexity of the phenomenon and the need for dimension reduction in relation to specific research or clinical problems is obvious. When meaningful summated measures of aspects of functioning and disability that might be used in the assessment of rehabilitation outcomes shall be designed, a sensible aggregation of items or, more generally speaking, variables along different dimensions of functioning is needed. Dimensions would then be represented by subscales compiled of items representing these dimensions. For each of the subscales a summary score might be calculated.

The WHOs International Classification of Functioning, Disability and Health [1] structures the complexity of human functioning and environmental factors by providing a list of 1424 categories of potential relevance ordered in four components and 34 chapters. This structure is mainly based on conceptual considerations. Previous research [7, 8] has shown that human functioning is indeed a multidimensional phenomenon and that the ICF structure is often compatible with empirical findings.

However, the dimensionality of the ICF has yet not been fully confirmed in empirical research. The ICF does not provide us with a guide on what categories or even items to select to best describe experiences of functioning in specific populations with the help of short and practicable measures. A classical approach to test the dimensionality of ICF data and to select variables would be the principal component analysis (PCA) [9]. A more innovative approach is graphical modelling developed in social network analysis, systems biology and genetics [10, 11, 12, 13]. This method is based on the estimation and visualisation of associations between variables such as ICF categories [7, 8]. The resulting association structure may then form different independent clusters of variables, i.e. dimensions. This may be interpreted as structural analogy to the PCA approach. In turn, the PCA results may be interpreted as a graphical model showing associations, i.e. edges, of variables to hidden nodes (the factors or latent dimensions).

The objective of this article is to examine whether and to what extent PCA and graphical models based on least average shrinkage selection operator (LASSO) regression differ in the dimension reduction of ICF data. The specific aims are to 1) introduce and compare PCA and graphical models and explain their potential for dimension reduction in human functioning, 2) to analyse the dimensionality of the ICF data in question with PCA and graphical models, 3) to compare the results generated with graphical models with the results of a PCA on the same data and 4) to discuss the consequences for dimension reduction in human functioning.

We use a data set of 1048 persons with spinal cord injury (SCI) from 14 different countries as a case in point.

4.2 Methods

4.2.1 Design

Secondary analysis of observational, cross-sectional data collected in the course of the development of ICF Core Sets for SCI in 14 countries [14] was performed.

All data collection in the original study was approved by the countries' ethical commissions and performed according to the principles of the Declaration of Helsinki.

4.2.2 Sample

The original sample was drawn by convenience. From June 2006 to January 2008, trained health professionals from 18 study centres in 14 countries collected data of 1048 adults with SCI due to an injury or disease with an acute onset [14].

4.2.3 Measurement

The functional problems of individuals with SCI were assessed using a case record form (CRF) comprising all 264 categories of the second level of the ICF. The CRF was completed by health professionals treating the patient in question. Information on the ICF components body functions and body structures was mostly based on the health professionals' assessment; information on the ICF component activities and participation was reported both by professionals and patients. The assessment of environmental factors was based on the patients' perception. The presence of a problem was binary coded for each category of body functions, body structures and activities and participation. Each category of environmental factors was recoded into two dummy variables, i.e. a facilitator and a barrier variable. Missing values were replaced using the AMELIA imputation technique [15].

4.2.4 Procedures for data analysis

To analyse the dimensionality of the ICF data, PCA and graphical models were applied.

4.2.5 Graphical models

A graphical model visualises conditional independencies among the ICF categories via a network of nodes and edges [10]. Nodes represent variables (the ICF categories). An edge between two nodes represents the existence of a conditional dependency between these two variables, even if controlling for all other variables. If two nodes are only connected through a third node (but not directly) then these two are conditionally independent given the third node. An algorithm to estimate such an association structure has recently been outlined by Strobl et al. [7]. The calculation of the association structures in the graphical model used here is based on LASSO regression [16]. The LASSO is a shrinkage and variable selection method for linear regression. In contrast to ordinary variable selection within regression modelling, it can handle very large numbers of variables (even beyond the number of patients). A so-called penalty term corrects for the sum of the absolute values of the coefficients in addition to the ordinary minimisation of the usual sum of squared errors. This ensures that models with a smaller number of predictors are favoured, if having a similar predictive power. Thus, the algorithm selects from the available set of variables only the most important ones. In the graphical model calculated here, a LASSO regression is performed for each ICF category with all the other ICF categories as predictors. Conditional dependency then corresponds to the non-zero coefficients in all of these models. To enhance stability of the model so-called bagging is applied, i.e. the algorithm is repeated for q different bootstrap samples [17]. Here, we chose q to be 500. The algorithm then counts in how many of the bootstrap samples a specific association showed up and considers only those associations that occurred in 100% of the samples. In the graphical model these highly stable associations are represented by edges between ICF categories. In the graph calculated for this article, we show only nodes with at least one edge. Furthermore, in a graphical model calculated with the LASSO procedure it can be adjusted for other potentially influential variables that shall not occur in the final model but be statistically held constant. Here, we adjusted for the patient's sex, situation (post-acute vs. chronic), age, years of formal education, level of SCI and completeness of injury.

The feasibility of graphical models for dimension reduction in human functioning has not yet been explored. However, a graphical model may produce several independent clusters of variables, i.e. ICF categories. These clusters form sub-graphs that are unconnected, i.e. conditionally independent, with other ICF categories in the graph. They may thus be interpreted as independent dimensions of human functioning.

4.2.6 Principal component analysis

Principal component analysis (PCA), in turn, is a well-established method for dimension reduction in the presence of categorical (here: binary) data [9, 18, 19]. A

PCA aggregates various variables into factors representing latent dimensions. On the basis of the covariance matrix of all interesting variables, factors are extracted according to predefined criteria, e.g. the number of resulting factors or the variance of the variables explained by one factor (eigenvalue). In the case of binary data, a covariance matrix is estimated based on the assumption of a multivariate binomial distribution of the analysed variables [18].

Each individual variable contributes to each factor in the form of factor loadings, i.e. the correlation of that variable with the overall factor. An initial factor solution can be orthogonally rotated in a space with as many dimensions as extracted factors. This procedure serves the minimisation of variance within factors and the maximisation of variance between factors (varimax rotation) and eases the interpretation of factors [9]. As a result of this procedure, each factor is independent of the other factors extracted, i.e. represents an independent dimension just as the sub-graphs. Factors can then be ordered by their eigenvalues, i.e. the variance of all variables in the model explained by the factor in question. A factor loading can, in turn, be interpreted as the association of one variable with a particular factor. Factor loadings are scaled from -1 to 1 so that large absolute values indicate more important variables for that particular factor, while values close to 0 indicate no relevance of this particular variable for the factor in question. Analogue to the graphical model explained earlier, we may discover several conditionally independent dimensions (factors). Each dimension is then represented by specific variables with high absolute loadings.

A PCA may also be interpreted and depicted as a graphical model showing edges of variables to hidden nodes (the factors). A variable loading on two different factors may be interpreted as a bridge between two factors or dimensions (Figure 4.2).

To determine the number of extracted factors with the help of an empirical criterion, the Kaiser criterion [20] is an often applied rule. The criterion states that an important factor should explain at least as much variance as a single standardised variable with variance 1 (i.e. factors with an eigenvalue equal to or above 1). Since in the study of graphical models we are particularly interested in the relationship of at least two variables, we tightened this criterion and only deemed factors important which had an eigenvalue of at least 2. Among these, we selected variables with an absolute factor loading exceeding 0.4 and matched each of them to the cluster with the highest overlap. The threshold of 0.4 was chosen because this represents a moderate to strong correlation in observational data given potential confounders that cannot be completely controlled by the design (e.g. randomisation is not possible). We compared this solution with another solution based on absolute factor loadings exceeding 0.7.

A commonly used alternative to the approach applied here is parallel analysis (PA). In this procedure, the eigenvalues of the factors extracted from the actual data set are compared with the eigenvalues of factors extracted from random permutations (of that data set). Eigenvalues are sorted in decreasing order. The number of relevant factors is then determined by the intersection of the consecutively decreasing eigenvalues of the actual and the permutated data sets [21].

Here, eigenvalues > 2 was preferred over PA to ease the comparison with the graphical model solution: As a graphical cluster comprises at least two variables, a factor should at least explain the variance of two variables. This criterion is more conservative then the PA. To test this assumption, we conducted a PA with our data. This yielded five additional factors. Moreover, on 1 factor only one variable had a loading higher than 0.4. However, the same was the case for our strategy though we had fewer factors. The overlap with the graphical model was marginally lower for the PA compared to our approach.

4.2.7 Comparison of procedures

There is a strong conceptual analogy between PCA and graphical modelling with regard to dimension reduction. Each of the procedures identifies subsets of strongly associated variables independent from the remaining variables. On the basis of this conceptual analogy, a comparison between the two procedures with regard to the emergent dimensionality of the data can be made. To compare the results of the graphical modelling with the results of the PCA the following approach was used here: In the final graph, we identified sets of variables (clusters) with a common association structure, i.e. all variables in the cluster are directly or indirectly connected with each other. As all variables in a cluster are independent of any variable not contained in that cluster, we expected to find factors with similar variable structures (recall that the individual factors are also independent of each other). More precisely, we analysed whether groups of variables with large absolute loadings on a particular factor could also be found as parts of a similar cluster within the graphical model. The procedure was the following: We determined the number of clusters matching factors by identifying for each ICF category in a particular cluster all factors which have at least one of these ICF categories with the cluster in common, i.e. have an absolute loading exceeding 0.4 (or 0.7). Likewise, for a particular factor we checked whether one or more of the ICF categories that had absolute loadings exceeding 0.4 (or 0.7) on that factor were also present in clusters of the graphical model. We define the cluster best matching a particular factor as the cluster that has the highest absolute number of ICF categories in common with the factor in question (this approach is illustrated in Figure 4.3). For example, for the factor 'facilitators in services, buildings and technology' 20 ICF categories with absolute factor loadings exceeding 0.4 were found within the cluster 'services, products and technology' making this cluster the one that is closest to the factor. Conversely, we define the best-matching factor of a particular cluster as the factor which has the highest absolute number of ICF categories (loading above 0.4 or 0.7) in common with the cluster in question. Besides overlap in absolute number of ICF categories assigned to the cluster or factor in question, we calculated relative overlap based on the average percentage of matching ICF categories from the perspective of the graphical model on one hand and the perspective of the PCA on the other. These percentages are the average of the ICF categories loading on the factor matched with any graphical cluster relative to the total number of items loading on the factor and the number of items in a graphical cluster that could be matched to any factor relative to the number of items in the cluster. The same procedure was applied to determine relative overlap for the best-matching clusters or factors only. To judge the overall overlap mean values, overall percentages calculated this way were calculated.

All statistical analyses were performed using R 2.9.0 [22] and SPSS 17 [23].

4.3 RESULTS

Table 4.1 – *Description of sample.*

Variable	Value	Mean(SD)/percent	Valid *N*
Age		42.2 (15.0)	1029
Sex	Female	22.5%	1048
Years of formal education		12.1 (4.5)	982
Time since injury (in years)		6.2 (9.9)	1029
Marital Status	Married	45.5%	1043
Paid employment	Yes	43.5%	1046
SCI Level	Tetraplegic	45.3%	1048
Impairment (ASIA)	Complete	52.2%	929
Situation	Chronic	53.4%	1048

The mean age of the patients was 42.2 (sd: 15.0): 22.5% were female, 45.3% had sustained a tetraplegia and 47.8% an incomplete SCI (Table 4.1). The countries differed with regard to the proportion of para- and tetraplegics: Australia (73.8%) reports the highest and Vietnam (24.4%) the lowest percentage of tetraplegics. Years of formal education ranged from 15.1 in New Zealand to 7.1 in Thailand.

Figure 4.1 shows the estimated graph that was calculated on the basis of all ICF categories assessed in the study. In total, 41 independent clusters or subgraphs can be identified (e.g. e460.b and e465.b: 'societal norms and attitudes (barriers)', in the topleft corner). Clusters vary in size from simple pairs up to a large entity of 33 ICF categories (top-middle to centre area, starting with e575.f: a cluster comprising environmental facilitators and barriers in 'services, products and technology'). Using the eigenvalue 42 criterion yields 24 factors when applying PCA. Figure 4.2 shows a graphical representation of this solution. On the basis of the variables sets with

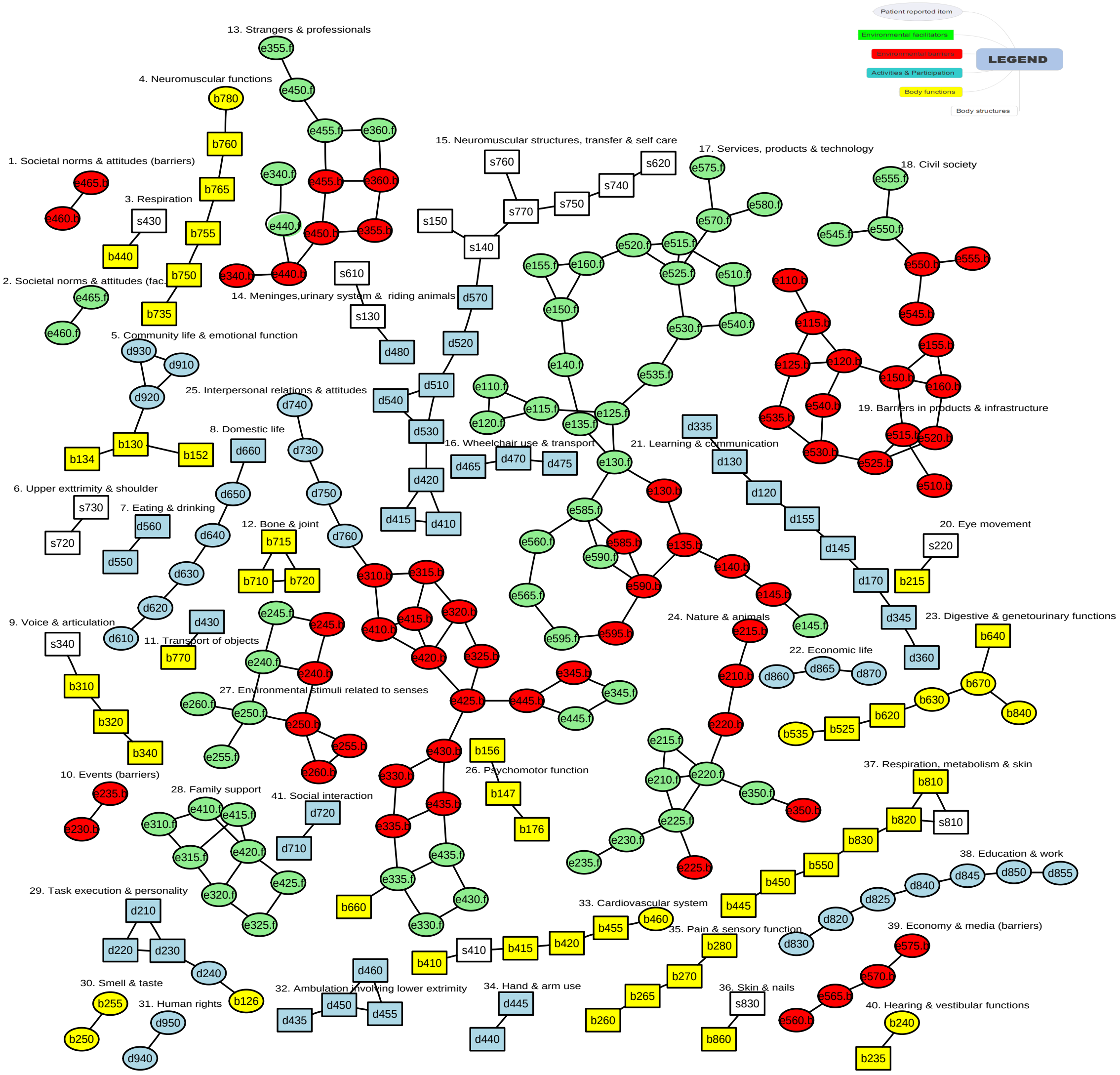

Figure 4.1 – *The resulting graph from the LASSO regression as outlined in the methods section. The colors represent the different ICF domains. Squares denote health professional-reported categories, whereas circles denote patient-reported categories. An edge between two ICF categories (represented as nodes) represents an association. Two nodes only connected through a third node are not directly associated but only via the latter.*

loadings greater than 0.4 the factors have been named. Similarly, we labeled the clusters found in the graphical model (Figure 4.1).

For variables with corresponding absolute loading exceeding 0.4 the overlap among these 24 factors with the corresponding matched clusters ranges from 9 to 100%. The average overlap is roughly 56%. In particular, several smaller clusters in the graphical model could not be matched with factors due to the, in total, larger number of clusters. In general, factors vary in size from 1 to 26 ICF categories with an average size of nine ICF categories. Within the best matching clusters we find

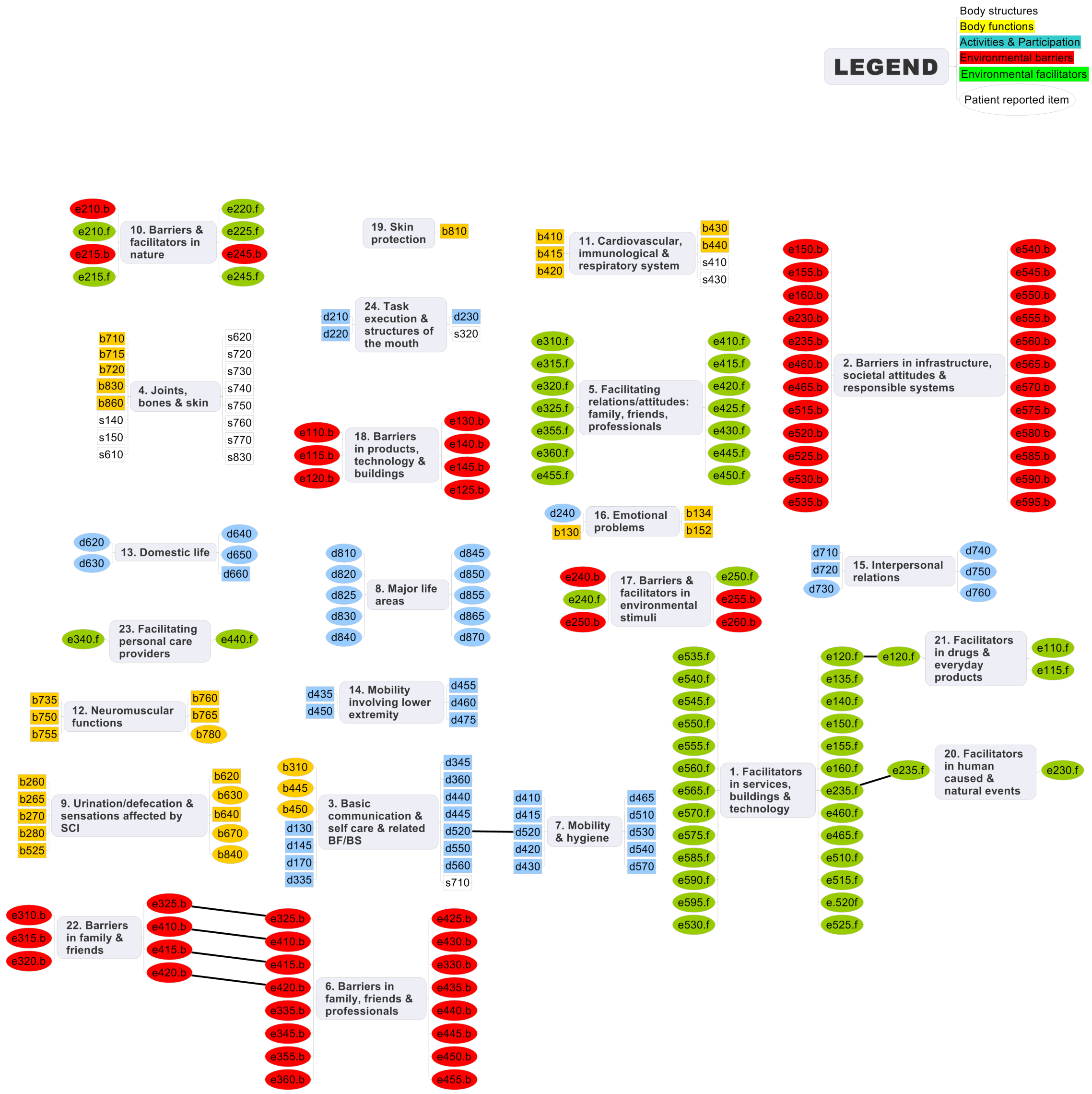

Figure 4.2 – *Graphical depiction of the factor solution found with PCA and orthogonal rotation. BF, body functions; BS, body structures. This visualises the structure of the factor loadings on each factor (thin lines) and among different factors (bold lines).*

on average six ICF categories that load on the corresponding factor. Thus, we find on average around 76% of a factor contained within the best-matched cluster. The total overlap of 56% is hence a result of the, on average, much larger size of the matched clusters relative to the average factor size. Details are given in Figure 4.3 and Table 4.2.

Table 4.2 – *Overlap between clusters generated with the graphical model and factors extracted with PCA.*

Factor (dimension)	No. of absolute loadings exceeding 0.4	No. of potential clusters per factor	No. of absolute loadings exceeding 0.4 in best matched cluster	Percent of all absolute loadings above 0.4 found in best-matching cluster	No. items in best matching cluster	Percent of absolute loadings above 0.4 in cluster relative to number of categories in cluster
1 Facilitators in services, buildings and technology	26	4	20	76.9	33	60.6
2 Barriers in infrastructure, societal attitudes and responsible systems	24	6	9	37.5	14	64.3
3 Basic communication and self-care and related body functions and structures	15	6	6	40.0	8	75.0
4 Joints, bones and skin	16	6	7	43.8	15	46.7
5 Facilitating relations with and attitudes of family, friends and professionals	14	3	8	57.1	8	100.0
6 Barriers in family, friends and professionals	16	2	11	68.8	25	44.0
7 Mobility and hygiene	10	2	8	80.0	15	53.3
8 Major life areas	10	2	7	70.0	7	100.0
9 Urination/defecation and sensations affected by SCI	10	2	6	60.0	7	85.7
10 Barriers and facilitators in nature	8	2	6	75.0	11	54.5
11 Cardiovascular, immunological and respiratory systems	7	2	4	57.1	6	66.7
12 Neuromuscular function	6	1	6	100.0	6	100.0
13 Domestic life	5	1	5	100.0	6	83.3
14 Mobility involving lower extremity	5	2	4	80.0	4	100.0
15 Interpersonal relations	6	2	4	66.7	25	16.0
16 Emotional problems	4	2	3	75.0	6	50.0
17 Barriers and facilitators in environmental stimuli	6	1	6	100.0	10	60.0
18 Barriers in products, technology and buildings	7	2	4	57.1	14	28.6
19 Skin protection	1	1	1	100.0	7	14.3
20 Facilitators in human caused and natural events	2	1	2	100.0	11	18.2
21 Facilitators in drugs and everyday products	3	1	3	100.0	33	9.1
22 Barriers in family and friends	7	1	7	100.0	25	28.0
23 Facilitating personal care providers	2	1	2	100.0	11	18.2
24 Task execution and structures of the mouth	4	1	3	75.0	5	60.0
Average over all 24 factors	8.9	2.25	5.9	75.8	13	55.7

Over 70% matches for percent of absolute loadings above 0.4 in cluster relative to number of categories in cluster are printed in italic.

In Figure 4.3, 54 potential combinations could be identified for loadings exceeding 0.4 and 11 for loadings exceeding 0.7. Averaging over the 54 nonzero percentages results in an overall overlap of 49.7%. Absolute factor loadings exceeding 0.7 result in an equivalent percentage of 59.3%. Restricting the comparison to the clusters best-matching each particular factor results in an overall overlap of 65.8% for absolute loadings exceeding 0.4 and of 58.2% if loadings above 0.7 are considered. Conversely, if we determine the factors best matching each cluster, the overall overlap is 56.4% for loadings above 0.4 and 59.3% for loadings exceeding 0.7. This corresponds to 36 and 11 clusters, respectively, which could be linked to a best-matching factor. For absolute loadings above 0.4, the clusters 13 and 18 had two bestmatching factors (both have the same number of matching items). For the calculation we thus averaged over all best-matching factors (i.e. for two clusters we average over two best-matching factors).

With both procedures we get environmental clusters or factors which are, if at all, only marginally (through one bridge) connected with functioning clusters or factors (bridges are represented as solid lines in Figure 4.2). In the graphical model, there is one activity and participation category, namely problems in family relationships (d760) that is conditionally dependent on barriers in family relations (e310b) and by one bridge connected to a larger structure comprising barriers in relations to family and friends, barriers in attitudes of these groups and barriers as well as facilitators in other groups with which daily interactions take place, e.g. colleagues or strangers (see cluster 25: 'interpersonal relations and attitudes'). Also, problems in procreation functions (b660) are connected with facilitators in people in subordinate positions (e335f) in that large cluster. In the PCA, no factor was found on which environmental factor and functioning variables jointly loaded. From 11 factors representing environmental dimensions in the PCA on only two barrier as well as facilitator variables loaded above 0.4: 'barriers and facilitators in nature' and 'barriers and facilitators in environmental stimuli'. The other environmental dimensions represent either barriers or facilitators only. In the graphical model, barriers and facilitators are indeed more often connected in one cluster, however within the clusters sub-structures representing barriers and such representing facilitators can be identified at one glance, e.g. in cluster 13 'strangers and professionals'. When we further look at double loadings of variables in the PCA and interpret them analogously to bridges between sub-structures, we find five connections of the different factors representing environmental categories. Again barriers are connected with barriers and facilitators with facilitators.

In the graphical model, we also often find separate clusters for body functions and structures on one hand, such as a cluster representing the 'cardiovascular system' (cluster 33), and separate clusters for activity and participation categories on the other, for instance two clusters representing problems in 'domestic life' (cluster 8)

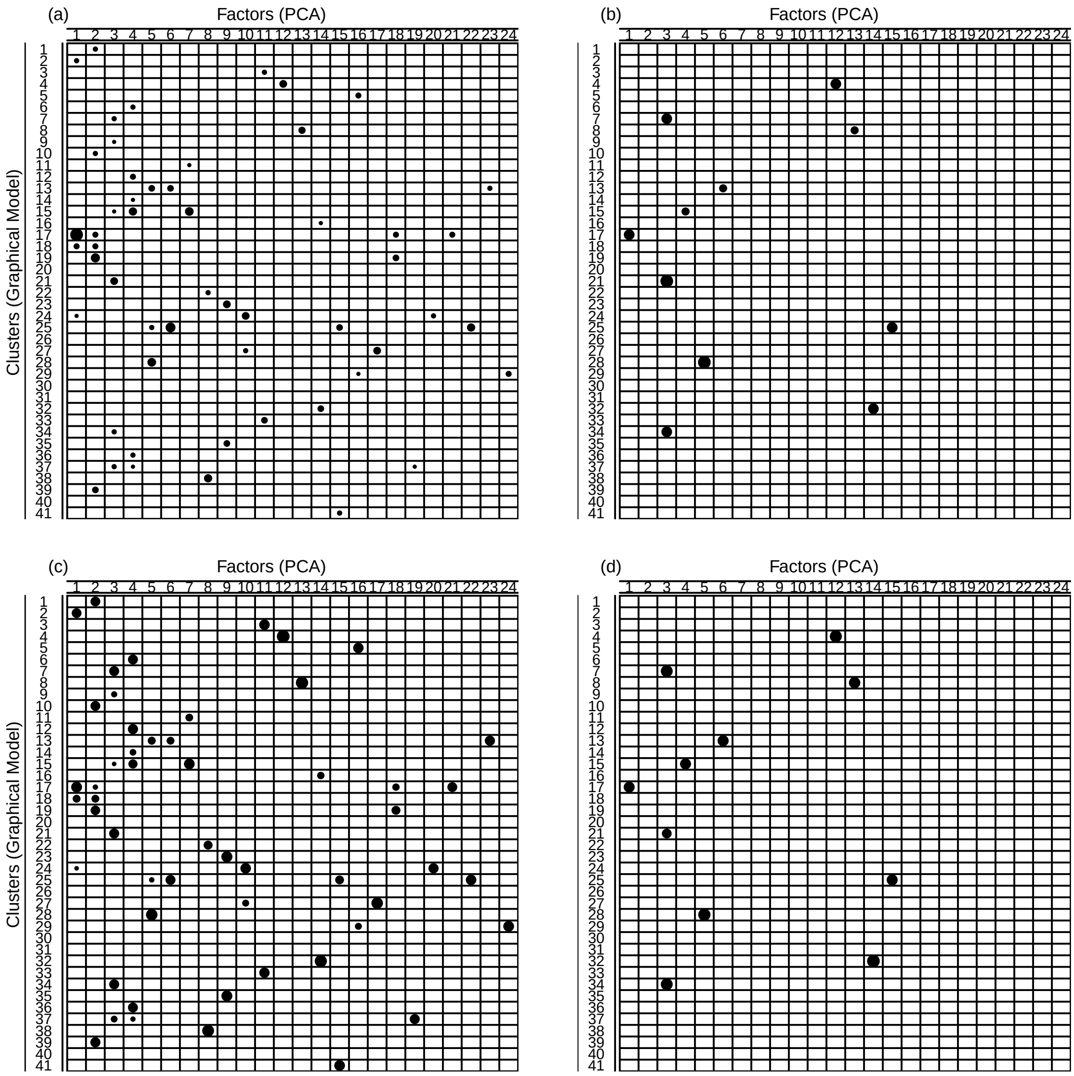

Figure 4.3 – *The overlap between clusters of the graphical model and factors of the PCA. A large circle represents a large total number of matched ICF categories. The best-matching cluster for a factor is represented by the largest circle within a column. Likewise, the best matching factor for a cluster is represented by the largest circle within a row. The absence of such a circle represents no overlap. The upper row shows the absolute overlap of clusters and factors for loadings exceeding a) 0.4 and b) 0.7. The lower row indicates the relative overlap in percent. These percentages are the average of the ICF categories loading on a particular factor that could be matched with a cluster relative to the total number of items loading on the factor and the number of items that could be matched to a factor relative to the total number of items in the cluster. Again, ICF categories were assigned to the factors based on absolute loadings exceeding c) 0.4 and d) 0.7. Figure 4.1 contains the labels for clusters 1 to 41. Figure 4.2 and Table 4.2 show the labels for factors 1-24.*

and in 'education and work' (cluster 38), respectively. Only four connections between activity and participation categories and body functions or structures exist in the graphical model within the clusters 5, 11, 14 and 15. The same holds true for the PCA results with body function and structure as well as activity and participation categories jointly loading on the three factors 3, 16 and 24. When we additionally consider that item d520 (caring for body parts) is loading on two factors, i.e. on factor 3 'basic communication, self-care and related body functions and structures' and factor 7 'mobility and hygiene', we see a comparable connection between body and activity and participation as in the graphical model.

In general, some clusters and factors discovered with the graphical model and the PCA, respectively mirror the same dimension with regard to the ICF content. For instance, entities representing problems in major life areas (factor 8) such as work and education (cluster 38) or in neuromuscular function (cluster 4 and factor 12) are found with both procedures. In other cases, the clusters of the graphs form subgroups of a PCA factor and vice versa. For instance, we find a factor representing problems in 'basic communication, self-care and related body functions and structures' (factor 3) and a cluster representing problems in 'learning and communication' (cluster 21) on one hand, and a cluster with problems in 'neuromuscular structures, transfer and self-care' (cluster 15) and a factor representing problems in 'mobility and hygiene' (factor 7) on the other.

4.4 Discussion

Both procedures applied illustrate the complexity of human functioning and disability in the mathematical sense in that not every element of a system can be connected with every other element. PCA and graphical models demonstrate the multidimensionality of functioning that needs to be considered when variables for the construction of meaningful outcome measures shall be selected. The construction of one summary score for an overall outcomes measure of functioning and disability is apparently not advisable.

The dimensionality of human functioning is not trivial. Although we have a great deal of overlap and dimensions that stay relatively stable across the two statistical procedures, we also find a different number of dimensions with each method and different compositions of dimensions. Most dimensions discovered with both procedures can be meaningfully interpreted from the clinical point of view, although some findings such as a connection between meninges, urological structures and riding animals for transportation (cluster 14 in Figure 4.1) are puzzling and may be statistical artefacts. With both methods, the ICF structure is quite well reproduced for activity and participation as well as environmental factors, while body functions from different ICF chapters are more often intermingled in the PCA factors. In particular, the connection of neuromuscular structures, transfer and self-care (cluster 15) as well as the similar connection of basic communication, self-care and structures of the neck (factor 3) show problems typical for the SCI population at hand.

Differences in the dimensional structure found with both models can also be due to adjustment for other influential variables such as level of SCI that is possible with graphical models but not with PCA. For example, in the PCA we find a factor relating problems in task execution with problems in structures of the mouth mirroring a problem typical for tetraplegics with high cervical lesions. When adjusting for level of SCI in the graphical model, we get a cluster relating the same problems in task execution with personality functions.

Since this is the first study to compare dimension reduction in human functioning with PCA and graphical models based on LASSO regression, there is no other study we could compare our findings with.

A limitation of the study at hand is the convenience sample with which the data were generated. We cannot exclude that the findings would be different in a random sample of the total population of persons with SCI from the countries examined. However, since national registries do not exist in most cases, such random sample is impossible to draw at the moment. Moreover, the limited number of associations between environmental factors and functioning categories may be due to the fact that environmental factors are patient reported while functioning categories are mostly based on health professional assessment. Indeed, differences of the self-reported and the health professionals' assessment need to be further analysed to grasp the full picture. These differences may even be seen as one way to depict outcomes in functioning [24]. To do this, however, 'subjective' and 'objective' assessments of the same ICF categories would be needed. This is not the case for the data in question.

Future research needs to compare PCAs' and graphical models' solutions for dimension reduction with other data so that step by step recurrent differences between the procedures will be discovered. A procedure to determine a common latent class structure of the two procedures could be to maximise overlap between the two methods. The overall relative overlaps from the graphical model and the PCA perspective can be modelled as functions of the number of factors, the maximum factor loadings and the number of clusters (i.e. the tuning parameter of the LASSO regression). Moreover, the stability of dimensions generated with both procedures needs to be tested across different settings such as higher- and lower-resourced countries [25], across the continuum of care and over the life span [26]. Also consequences for variable selection need to be elaborated. A graphical model does, for instance, not only show common associations but also the structure of these associations which is an important advantage over the PCA. A chain of categories may have different consequences for variable selection in that dimension than a triangle structure. In the latter case, only one variable may be selected while in the former more might be relevant to avoid losing information.

Functioning is a multidimensional, complex and dynamic outcome. We thus recommend being extremely careful with dimension reduction and corresponding item aggregation in outcomes research based on statistical procedures alone. Theoretical considerations and clinical significance should always guide statistics aiming at the reduction of dimensions: every good model starts from a research question relating to particular levels of analysis, i.e. bulldozers should not be modelled with quarks [27]. In the light of the obvious lack of sophisticated theory in functioning and rehabilitation, qualitative research and metasynthesis of qualitative studies may be starting point for the development of theoretical assumptions about dimensions [28]. Furthermore, the stability of emergent dimensions of ICF data which are compatible with theoretical and clinical considerations as well as consumers' experience is a most important point in advancing better outcome measures of functioning and disability. We recommend not choosing one but at least two different statistical procedures for dimension reduction, for instance PCA and LASSO regression and look at the dimensions that stay stable with both procedures. Then subgroup analysis should be performed to test dimensional stability across different populations and settings.

Finally, the lived experience of persons with disabilities is always more complex than dimensions we extract by theoretical or statistical means. To properly reduce complexity without losing the connection to this lived experience is what outcomes research in disability is all about.

Acknowledgements

We thank Sibylle Graf for her help with the manuscript. Markus Kalisch and Peter Bühlmann from the Swiss Federal Institute of Technology in Zurich provided valuable feedback in discussions. We thank Will Taylor for his highly stimulating comments on the paper.

Literaturverzeichnis

[1] WHO. International Classification of Functioning, Disability and Health: ICF, 2001.

[2] M. Leonardi, J. Bickenbach, T. B. Ustun, N. Kostanjsek, and S. Chatterji. The definition of disability: what is in a name? *Lancet*, 368(9543):1219–21, 2006.

[3] J. E. Bickenbach, S. Chatterji, E. M. Badley, and T. B. Ustun. Models of disablement, universalism and the international classification of impairments, disabilities and handicaps. *Soc Sci Med*, 48(9):1173–87, 1999.

[4] S Nagi. Disability concepts revisited: implications for prevention. In A Pope and A Tarlov, editors, *Disability in America: a national agenda for prevention*, pages 309–327. National Academy Press, Washingtion, 1991.

[5] N Cantor and CA Sanderson. Life task participation and wellbeing: the importance of taking part in daily life. In D Kahnemann, E Diener, and N Schwarz, editors, *Well-being: the foundations of hedonic psychology*, pages 230–243. Russel Sage Foundation, New York, 1999.

[6] I. B. Wilson and P. D. Cleary. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*, 273(1):59–65, 1995.

[7] R Strobl, G Stucki, E Grill, M Müller, and U Mansmann. Graphical models illustrated complex associations between variables describing human functioning. *J Clin Epidemiol*, (62):922–33, 2009.

[8] J. D. Reinhardt, U. Mansmann, B. A. Fellinghauer, R. Strobl, E. Grill, E. von Elm, and G. Stucki. Functioning and disability in people living with spinal cord injury in high- and low-resourced countries: a comparative analysis of 14 countries. *Int J Public Health*, 56(3):341–52, 2011.

[9] G Duntemann. *Principal components analysis*. Sage, New York, 1989.

[10] N. Friedman. Inferring cellular networks using probabilistic graphical models. *Science*, 303(5659):799–805, 2004.

[11] L. Hood, J. R. Heath, M. E. Phelps, and B. Lin. Systems biology and new technologies enable predictive and preventative medicine. *Science*, 306(5696):640–3, 2004.

[12] S Wasserman and K Faust. *Social network analysis: methods and applications*. Cambridge University Press, Cambridge, 1994.

[13] K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. L. Barabasi. The human disease network. *Proc Natl Acad Sci U S A*, 104(21):8685–90, 2007.

[14] F. Biering-Sorensen, M. Scheuringer, M. Baumberger, S. W. Charlifue, M. W. Post, F. Montero, N. Kostanjsek, and G. Stucki. Developing core sets for persons with spinal cord injuries based on the International Classification of Functioning, Disability and Health as a way to specify functioning. *Spinal Cord*, 44(9):541–6, 2006.

[15] G. King, J. Honaker, A. Joseph, and K. Scheve. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, 95(1):49–69, 2001.

[16] Nicolai Meinshausen and Peter Buehlmann. High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*, 34:1436–1462, 2006.

[17] L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.

[18] J Jackson. *A user's guide to principal components analysis*. Wiley, New York, 2003.

[19] IT Jolliffe. *Principal component analysis*. Springer, New York, 2002.

[20] H. F. Kaiser. The application of electronic computers to factor analysis. *Educ Psychol Meas*, 20:141–151, 1960.

[21] B. P. O'Connor. SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. *Behav Res Methods Instrum Comput*, 32(3):396–402, 2000.

[22] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0.

[23] N. N. *SPSS for Windows. Rel. 17.0.0.* SPSS Inc., Chicago, 2008.

[24] N. M. Kayes and K. M. McPherson. Measuring what matters: does 'objectivity' mean good science? *Disabil Rehabil*, 32(12):1011–9, 2010.

[25] JD Reinhardt, BA Fellinghauer, R Strobl, and Stucki G. Dimension reduction in human functioning and disability outcomes research: graphical models versus principal components analysis. *Disability and rehabilitation*, (32):1000–10, 2010.

[26] G. Stucki, N. Kostanjsek, B. Ustun, and A. Cieza. ICF-based classification and measurement of functioning. *Eur J Phys Rehabil Med*, 44(3):315–28, 2008.

[27] N. Goldenfeld and L. P. Kadanoff. Simple lessons from complexity. *Science*, 284(5411):87–9, 1999.

[28] W. M. Levack, N. M. Kayes, and J. K. Fadyl. Experience of recovery and outcome following traumatic brain injury: a metasynthesis of qualitative research. *Disabil Rehabil*, 32(12):986–99, 2010.

KAPITEL 5

GRAPHICAL MODELING CAN BE USED TO ILLUSTRATE ASSOCIATIONS BETWEEN VARIABLES DESCRIBING FUNCTIONING IN HEAD AND NECK CANCER PATIENTS

Sven Becker[a], Ralf Strobl[b], Alarcos Cieza[b], Eva Grill[b], Ulrich Harreus[c], Uta Tschiesner[a]

[a]*Department of Otorhinolaryngology, Head and Neck Surgery, Ludwig-Maximilians-Universität München, Munich, Germany*

[b]*ICF Research Branch of the WHO Collaborating Center for the Family of International Classifications at the German Institute of Medical Documentation and Information (DIMDI), Institute for Health and Rehabilitation Sciences, Ludwig-Maximilians-Universität München, Munich, Germany*

[c]*Department of Otorhinolaryngology, Head and Neck Surgery, Friedrich-Alexander-University, Erlangen, Germany*

This Chapter was published as

Becker S, Strobl R, Cieza A, Grill E, Harreus U, Tschiesner U: **Graphical modeling can be used to illustrate associations between variables describing functioning in head and neck cancer patients.** *J Clin Epidemiol* 2011, **64**:885-92.

Abstract

Objective: To examine the associations between variables of functioning measured by the International Classification of Functioning, Disability and Health (ICF) in head and neck cancer (HNC) patients by means of graphical modeling.

Study Design and Setting: Graphical modeling was used on a data set of a cross-sectional multicentric study of 145 patients with HNC. Functioning was qualified using the extended ICF checklist. Multiple imputation was used to handle missing data. The least absolute shrinkage and selection operator for generalized linear models was used to identify conditional associations between the ICF categories. Bootstrap aggregating was used to enhance the accuracy and validity of model selection.

Results: The resulting graph shows largely meaningful associations between the ICF categories. One central point could be visualized consisting of a circular path of *d330 Speaking*, *d350 Conversation*, *b510 Ingestion functions*, *s320 Structure of mouth*, and *b310 Voice functions*. Another important structure in the graph were the bow-shaped associations beginning with *d335 Producing nonverbal messages* to *b130 Energy and drive functions*.

Conclusion: Graphical modeling can be used to describe associations between different areas of functioning in HNC patients. They found associations can be the basis for improved rehabilitation and gives a deeper understanding of functioning in HNC patients.

Keywords: *Tumor; Head and neck cancer; Graphical models; Functioning; International Classification of Functioning, Disability and Health (ICF); Rehabilitation*

5.1 Introduction

A decrease in functioning is frequently seen in patients after initial therapy of head and neck cancer (HNC) [1, 2, 3]. Impairments in body function and body structure including eating, breathing, speech production, and pain can lead to changes in self-image, worsened relationship with the partner, and increased social isolation [4, 5]. The result is an enormous loss in quality of life.

Adequate rehabilitation is therefore central for patients with HNC as they have a good survival rate. To optimize interventions aimed at maintaining functioning and minimizing disability, a proper understanding of the patients' functioning and health status is needed [6]. The International Classification of Functioning, Disability, and Health (ICF) endorsed by the World Health Organization in 2001 provides a useful framework for classifying the components of health and consequences of a disease. It aims at providing a unified language for the description of health conditions in rehabilitation. According to the ICF, the consequences of a disease may concern body functions, body structures and the performance of activities and participation in life situations. Health states and the development of disability are modified by contextual factors such as environmental and personal factors [7].

It is well known that patients after initial therapy of HNC suffer from a variety of problems in functioning because of the treatment and the cancer itself. Surgical procedures to the larynx can impair breathing and voice production [8, 9]. Furthermore, extensive surgical procedures to the head and neck region can cause disfiguring changes to the appearance of a patient [10]. Postoperative and especially postradiation swelling can lead to problems in swallowing in the early phase of rehabilitation, whereas xerostomia because of the damage of the salivary glands causes swallowing problems in the long-term follow-up of these patients [8, 11]. Another main topic in the rehabilitation of HNC patients is the control of pain because of scarring, mucositis, and the hardening of head and neck musculature [11]. These impairments can entail occupational retirement and can cause psychological problems such as depressive disorders, which not only affect the patient but also influence the relationships to the family and the partner [12, 13]. In contrast, little is known about how these different parts of functioning are connected with each other or how they interact. So far it is not comprehensively explored which areas of functioning are conditionally dependent and what variables can function as confounders. Furthermore, no efforts were made to visualize these associations. To improve rehabilitation of HNC patients, a deeper understanding of the dependencies and interactions between these different areas of functioning is needed. One possibility to visualize possible associations could be graphical modeling. The graphical description of the different associations can direct the focus to important central points in the graph. In general, interesting geometrical structures can give indications how to analyze data [14]. Simple structures, for example, can be crossings of different arms in the graph with one category in the center suggesting its importance. Moreover, circular paths can illustrate the high interdependency of a number of categories. Theoretically, circular paths in graphical models can contain variables, which are sufficiently correlated to form a common dimension for measurement [15]. These central points can be the target for interventions improving functioning in the future and can also help to optimize given resources.

With the ICF, it is possible to analyze the association of elements of functioning on the level of single categories using graphical modeling. The model was developed by Strobl et al. [14] to illustrate associations between variables describing human functioning when using the ICF as a reference. This graphical model forms a comprehensive probabilistic tool to analyze and visualize associations between different ICF categories impaired in HNC patients.

Therefore, the objective of this study was to illustrate associations between variables in functioning in HNC patients using the ICF as a reference. Furthermore, the complex associations between these variables should be examined to obtain important targets to improve the rehabilitation of HNC patients in the future.

5.2 Methods

Data from a cross-sectional study with a convenience sample of 145 patients with HNC was used for graphical modeling [16]. The study took place in nine study centers in Greece, Germany, Poland and Turkey and was part of the development of ICF Core Sets for Head and Neck Cancer. Study centers were found by a worldwide public invitation to participate in this development process [17]. In each study center, one health professional familiar with the ICF was responsible for the data collection. The study protocol and consent forms were approved by the Ethics Committee of the University in Munich, Germany and the Ethic Committee of each of the study centers.

5.2.1 Extended ICF checklist

Health professionals were asked to complete the 'extended ICF checklist for HNC' for each of their HNC patients. Inclusion criteria were as follows: (1) cancer of at least one of the following cancer locations: oral region, salivary glands, oropharynx, hypopharynx, or larynx; (2) patients must have completed initial cancer treatment; (3) at least 18 years of age; (4) comprehension of the purpose of the study; and (5) giving signed informed consent. The extended checklist is based on the generic ICF checklist from WHO (130 categories) [18] and was amended by additional HNC-specific categories that were identified in a content comparison between the ICF- and HNC-specific validated questionnaires (14 categories) [19]. All together 144 categories were referred to as the 'extended ICF checklist for HNC.'

The level of impairment or restriction was assessed according to the ICF qualifier scale: 0, no impairment/restriction; 1, mild impairment/restriction; 2, moderate impairment/ restriction; 3, severe no impairment/restriction; 4, complete no impairment/restriction. For the environmental factor component, each category can be either a facilitator or a barrier to the patient. To denote that the category is a facilitator, a positive sign is added (e.g., +3) and to denote the category as a barrier, a negative sign is added (e.g., -3). The option 'not specified' (ns) is applied when the available information is not sufficient to quantify the severity of the problem, and the option 'not applicable' (na) when the category is not applicable to the patient. For impairments caused by comorbidity and not by HNC, the option (C) 'comorbidity' was filled in.

5.2.2 Data collection

Data collection and patient interviews were performed by health professionals trained in the ICF at each of the study centers. Patient selection was done with consecutive sampling. The training for the health professionals involved familiarization with the WHO model of functioning and disability and with the ICF. Detailed and precise guidelines to perform the structured interviews were provided by the coordinating center in Munich, Germany. Completion of the extended ICF checklist for HNC was based on structured patient interview, professional observation, and the medical records.

5.2.3 Analysis

To characterize the study population, we used mean and standard deviation for continuous data and frequency for categorical data. To be able to perform graphical modeling on the data set of the extended ICF checklist, data was converted into a binary form. No impairment/restriction, ICF qualifier 0 was converted into 0 (no impairment) for the graphical modeling. Mild, moderate, severe, or complete impairment/ restriction, ICF qualifier 1, 2, 3, or 4 were merged into 1 (impairment) for further analysis. Positive or negative signs in the component environmental factors were not considered. Categories that were filled in as comorbidities (C), 'not applicable' (na), or 'not specified' (ns) were coded as missing.

As the sample size was rather small, it was advisable to reduce the number of variables first. Two health professionals experienced in the treatment of HNC patients decided on which ICF categories should be the basis for the calculation of the model. Both experts were involved in former studies concerning the development of the ICF Core Sets for HNC and were familiar to the ICF [19, 20, 21]. The aim of the selection was to minimize the categories on the one hand to optimize the calculation of the model, and to include at least three or more categories from each of the four different ICF components to get a preferably comprehensive description of the problems HNC patients after initial therapy are faced with. These categories should nicely describe the complicated interaction. This reduced the number of falsely included edges in the final model and enhanced interpretability. We preferred the expert opinion as experts understand the concept of causality in a broader sense, then an automatic method would do. The resulting 34 second-level categories, 12 from the component body functions (b), 3 from the component body structure (s), 15 from the component activity and participation (d), and another 4 categories from the component environmental factors (e) can be seen in Table 5.1.

Missing values in the data set may introduce bias and can lead to false conclusions under certain circumstances. Therefore, we used multiple imputation to handle missing data [22]. This method was shown to be superior regarding variance in simulation and real-data settings to complete case analysis and showed similar results to other common missing value approaches [23, 24]. For imputation, we used the Multivariate Imputation by Chained Equations package running under R [25, 26].

5.2.4 Graphical modeling

Graphical models allow displaying the net-like dependence structures by drawing a graph describing the conditional associations between the variables. Thus, a graphical model is a graph with nodes representing random variables (e.g., *d350 Conversation*) and edges between two nodes representing a conditional association. The edges used in graphical models to connect two variables indicate that the two variables are dependent even if controlling for all other variables. It is important to clarify the meaning of conditional dependency. Firstly, conditional dependency can also be observed in the presence of unconditional independency. Furthermore, it should not be mixed with causality as it is a necessary but not sufficient condition for causality. Strobl et al. [14] could show that graphical modeling can be adapted to illustrate complex associations between variables of human functioning when using the ICF as a reference.

To give an example, the chain *d330 Speaking - d350 Conversation - d360 Using communication devices and techniques* shows the dependency structure of three variables describing aspects of functioning in HNC cancer patients (see Fig. 5.1). Thus, information about restrictions in *Speaking* do not give any further information about restrictions in *Using communication devices and techniques*, once we know restrictions in *Communication*. So, if the variables are conditionally independent there will be no direct edge between them. In this example, *d330* and *d360* are conditionally independent because they are not connected by an edge when *d350* is given. In addition, *d350* can be seen as a confounder variable regarding the relation of *d330* and *d360*. For further investigation of the association between *d330* and *d360*, we should control for *d350*.

Table 5.1 – *Short description of the ICF categories used for the graphs*

ICF Code	ICF Code description
	ICF component body function
b130	Energy and drive functions
b280	Sensation of pain
b310	Voice functions
b320	Articulation functions
b340	Alternative vocalization functions
b440	Respiration functions
b450	Additional respiratory functions
b460	Sensations associated with cardiovascular and respiratory functions
b510	Ingestion functions
b515	Digestive functions
b530	*Weight maintenance function*
b710	Mobility of joint functions
	ICF component body structure
s320	Structure of mouth
s430	Structure of respiratory system
s710	Structure of head and neck region
	ICF component activity and participation
d175	*Solving problems*
d310	Communicating with - receiving - spoken messages
d315	Communicating with - receiving - nonverbal messages
d330	Speaking
d335	Producing nonverbal messages
d350	Conversation
d360	Using communication devices and techniques
d550	Eating
d560	Drinking
d570	Looking after one´s health
d720	Complex interpersonal interaction
d760	Family relationship
d770	Intimate relationship
d850	*Remunerative employment*
d920	Recreation and leisure
	ICF component environmental factors
e125	Products and technology for communication
e225	Climate
e310	*Immediate family*
e580	Health services, systems and policies

Abbreviation: ICF, International Classification of Functioning, Disability and Health.
Categories without any association to one of the other 33 categories are written in *italics*.

The associations between different ICF categories were identified by means of a penalized logistic regression model, so-called least absolute shrinkage and selection operator [27]. The visualization of the graph was done using the R package Rgraphviz [25, 28] using an automatic algorithm for drawing undirected graphs [29].

We added the odds ratios to the edges of our graph estimated from the penalized logistic regression model using OR5exp(regression coefficient). Higher odds ratios indicate stronger positive associations. An odds ratio of 1 would indicate no association and an odds ratio smaller than 1 would indicate a negative association.

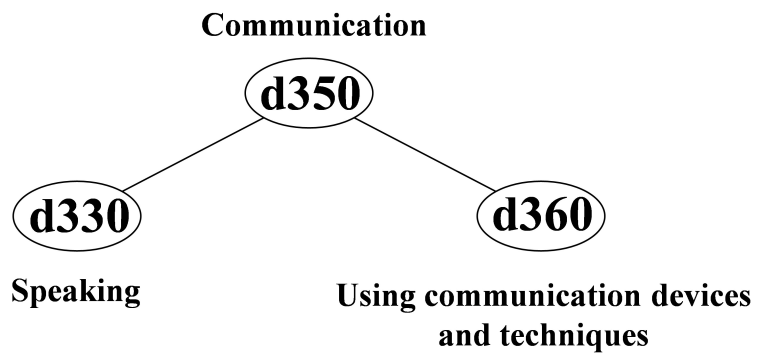

Figure 5.1 – *Simple graphical model of the associations between three variables of functioning in head and neck cancer patients.*

The thickness of the edges corresponds to the strength of the association using the following rule: the 20% with the strongest associations are printed in bold and the 20% with the weakest associations are in gray. Negative associations are plotted with dashed lines.

5.3 RESULTS

All together 145 patients with HNC were interviewed between April and October 2007 in nine study centers in four European countries. There were 108 patients interviewed in Germany, 20 in Poland, 11 in Greece, and 6 in Turkey. The sociodemographic and tumor-specific characteristics of the study population are presented in Table 2. Most participants were male (74%). Patients had a mean age of 59 years ranging from 31 to 80 years. Sixty-one (42%) of the patients had just an operation, 72 (50%) had an operation plus radiotherapy, and another 12 (8%) underwent radio(chemo)therapy before entering the study. There were 122 complete data sets beside 23 incomplete data sets with at least one missing statement.

Category *b530 Weight maintenance function*, *d175 Solving problems*, *d850 Remunerative employment* and *e310 Immediate family* had no associations with any other category used for the calculation of the model. Each of them represents an isolated node without any edge to one of the 33 further categories in the graph. Therefore, they are not shown in Figs. 5.2 and 5.3.

In addition, most of the categories showed meaningful associations with other categories (see Fig. 5.2). There were interesting circle-like associations between *d330 Speaking*, *b310 Voice functions*, *s320 Structure of mouth*, *b510 Ingestion functions*, and *d350 Conversation*. On the basis of this circle, there were further interesting associations to other categories. *d360 Using communication devices and techniques*, connected with *d350 Conversation* had further associations to *e125 Products and technologies for communication* and *d920 Recreation and leisure*, which was further connected to *d770 Intimate relationships* (see Figs. 5.2 and 5.3, upper right part).

Table 5.2 – *Sociodemographic and tumor-specific characteristics of the study population*

	Patient demographics and tumor characteristics	
	No. of patients	(%)
All patients	145	100
Mean age, yr (range)	59 (31-80)	
Gender		
Male	108	74
Female	37	26
Study centers		
Germany	108	74
Poland	20	14
Greece	11	8
Turkey	6	4
Complete observations		
Complete data sets	122	84
Incomplete data sets	23	16
Tumor location		
Oral cavity	55	38
Oropharynx	42	29
Hypopharynx	22	15
Larynx	50	34
Salivary glands	8	6
> 1 location	23	17
Staging		
Stages I and II	46	31
Stages III and IV	92	64
Not available	7	5
Treatment		
Operation (OP)	61	42
OP and radio(chemo)therapy	72	50
Radio(chemo)therapy	12	8

d330 Speaking had a further association to *b320 Articulation functions*. The category *b310 Voice function* had a further association with *b340 Alternative vocalization functions* and *s320 Structure of mouth* had a meaningful connection to *d560 Drinking*, which was further connected to *d550 Eating* (see Figs. 5.2 and 5.3, lower middle part). The last category in this circle *b510 Ingestion functions* had an association to *b280 Sensation of pain*, which had a further negative association with *s710 Structure of the head and neck region* (see Fig. 5.2, middle part and Fig. 5.3, lower right part).

A further bow-like association could be visualized between the categories *d335 Producing nonverbal messages, d315 Communicating with - receiving - nonverbal messages, d310 Communicating with - receiving - spoken messages, d720 Complex*

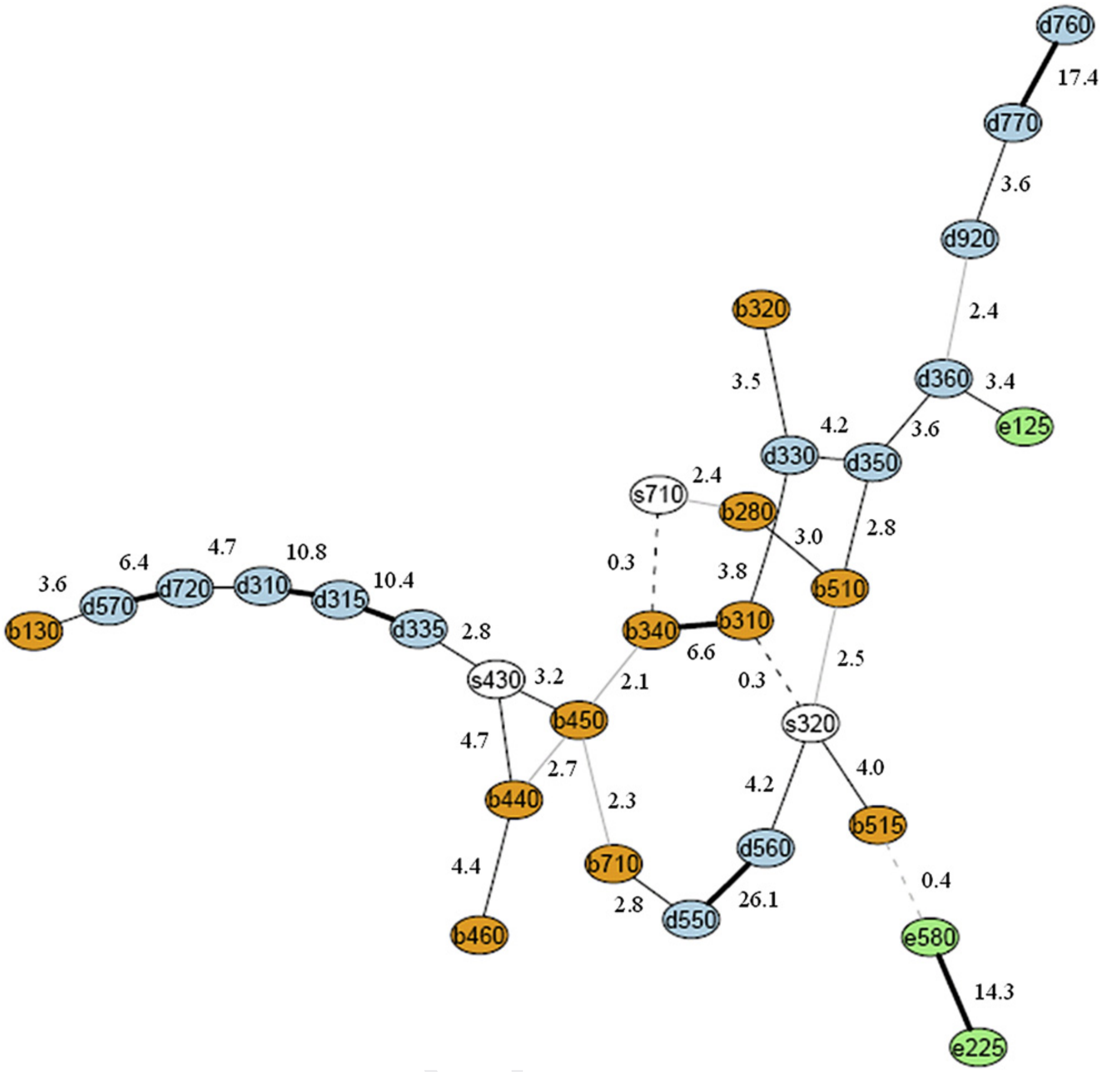

Figure 5.2 – *Graphical modeling of the associations between the different International Classification of Functioning, Disability and Health (ICF) categories. ICF categories from the component body function (orange), body structure (white), activities and participation (blue), and environmental factors (green). Please, find the full descriptions of the ICF categories in Table 1. Each edge is labeled with the odds ratio of the association between the two variables estimated from the penalized logistic regression models using OR5exp(regression coefficient). The presentation of the edges corresponds to the strength of association, the 20% with the strongest associations are printed in bold and the 20% with the weakest associations are printed in gray. Negative associations are plotted with dashed lines.*

interpersonal interaction, d570 Looking after one's health and *b130 Energy and drive functions* (see Fig. 5.2, left part).

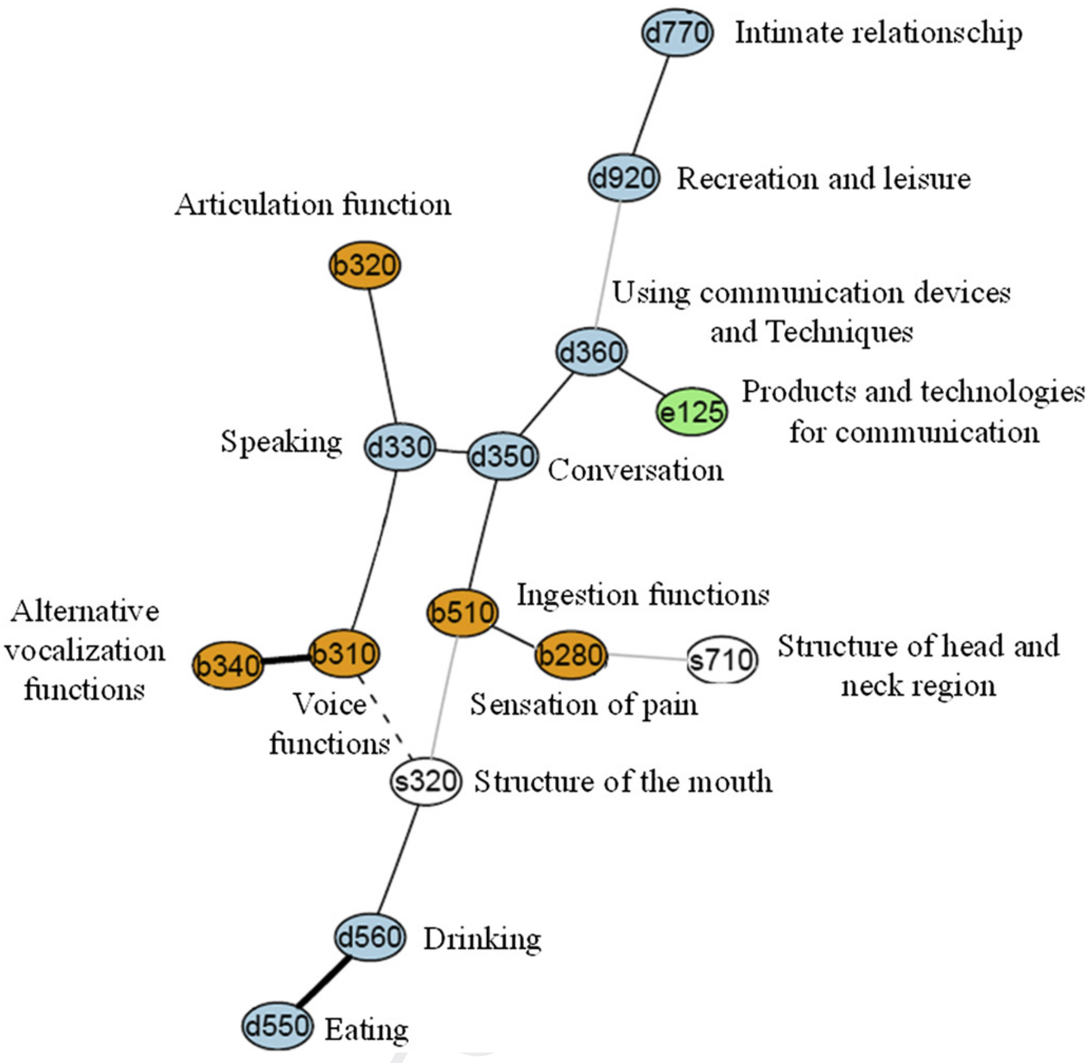

Figure 5.3 – *Meaningful associations based on a circle out of d330 Speaking, d350 Conversation, b510 Ingestion functions, s320 Structure of mouth, and b310 Voice function. The presentation of the edges corresponds to the strength of association, the 20% with the strongest associations are printed in bold and the 20% with the weakest associations are printed in gray. Negative associations are printed in dashed lines.*

5.4 Discussion

This is the first study using graphical modeling to illustrate associations between variables describing functioning in HNC patients based on the ICF. Therefore, we used a graphical model established by Strobl et al. [14] on a data set of an empirical cross-sectional study with a convenience sample of 145 HNC patients [16]. Although the results showed mostly meaningful associations between the different second-level ICF categories, there are a few points that should be discussed when using this model.

The first point that can lead to bias in the model is the data collection procedure. The health professional who carried out the interview may have introduced artificial associations when interpreting the answers of the patient. We tried to minimize that problem by training and supervising the interviewers [16].

Secondly, to handle the missing data we applied multiple imputation what has been shown to be superior to other imputation methods [30, 31, 32]. Anyhow, multiple imputation can lead to artificial associations.

Thirdly, although nine study centers have worked together in the data collection process, most of the patients came from Germany. However, when comparing results from the different study centers, there was no systematic difference between German and non-German study centers [16]. Still, the sample size of 145 patients is relatively small and the sociocultural context of European countries might differ from that in American, Asian, or African countries.

Fourthly, data from patients of different treatment regimes were pooled together. Previous studies described different impact on health-related quality of life and functioning for radiation and radio(chemo)therapy compared with surgery alone [33, 34, 35]. In our study, 133 patients were treated by surgery alone or in combination with radio(chemo)therapy beside 12 patients, which were treated by radio(chemo)therapy alone. No significant differences in symptoms for these two treatment groups could be shown in a previous study for the data [16]. Therefore, we decided that pooling of the two treatment groups is justified.

Anyhow, the results of the study show many clinically meaningful associations between different variables of functioning in HNC patients what gives evidence to the high face validity of the used method. HNC patients after initial therapy have to deal with a broad variety of impairments in different fields of functioning. Our results can be used to understand the interaction of different fields of functioning in these patients and can lead to an improvement in the rehabilitation of HNC patients. Interesting associations were seen around the circular path consisting of *d330 Speaking, d350 Conversation, b510 Ingestion functions, s320 Structure of mouth*, and *b310 Voice functions* (see Figs. 5.2 and 5.3 middle part). This circle illustrates the high interdependency of these five categories and is concordant with the literature [9, 36]. One of the main topics in rehabilitation is the restoration of voice (*b310*) what gives the patient the possibility to communicate with his environment. This can be done by using different communication devices (*d360*) such as electronic speechgenerating devices or voice prostheses after laryngectomy (*e125*) [9, 37, 38] or just by learning how to speak again with the new anatomical conditions after cancer therapy (*d330*) [36]. The newly achieved possibility to have a conversation (*d350*) will help socializing (*d920*) and can improve the relationship with the partner (d770) and the family (*d760*) [39, 40, 41] (see Figs. 5.2 and 5.3, upper right part).

Another main area in the rehabilitation of HNC patients is the recovery of oral food and liquid intake [42]. The category *b510 Ingestion functions* includes a variety of functions that can be impaired after initial tumor therapy including chewing and biting, the manipulation of food in the mouth, salivation, and swallowing. Patients after extensive surgery, for example, for the tongue (*s320*) or the larynx often suffer from dysphagia, pain or aspiration of food (*b510*), whereas patients after additional or primary radiation therapy more often have problems in salivation and therewith connected teeth problems [43]. Both groups can be affected by pain (*b280*) because of epithelial damage of the aero-digestive tract. Furthermore, pain can occur in different structures of the head and neck region (*s710*) including joints, muscles, and ligaments (see Fig. 5.2, middle part and Fig. 5.3, lower right part). Scaring or postradiation changes can lead to this pain [11]. Especially, pain or movement restrictions in the temporomandibular joint (*b710*) can be problematic when trying to eat (*d550*) or drink (*d560*) [44]. Therefore, big efforts are and were made to preserve these functions by reconstructing defects, for example, with free tissue transfer [45, 46] or to improve radiation procedures and protocols with different additional agents [47].

The negative association between *b310 Voice functions* including the production of various sounds by the passage of air through the larynx and *s320 Structure of mouth* can be explained by the different tumor locations. Thus, an alteration in the structure of mouth (*s320*), for example, after surgical replacement of a part of the tongue because of a carcinoma of the tongue does exclude patients with a carcinoma of the larynx as the occurrence of carcinomas in two locations was rarely seen in these patients. Therefore, an impairment of *s320* does not directly lead to an impairment in voice production (*b310*) and vice versa.

The bow-shaped associations beginning with *d335 Producing nonverbal messages* to *b130 Energy and drive functions* (Fig. 5.2, left part) show how important communication from the early posttreatment situation is for a good overall rehabilitation. Even if the production (*d335*) and communication with nonverbal messages (*d315*) is a problem in the early stage, for example, after laryngectomy, it is the first step to start interpersonal interaction after initial treatment. After the rehabilitation of voice and speech, a more complex interpersonal interaction (*d720*) will be possible by communicating with spoken messages (*d310*). This newly acquired possibility to communicate with relatives on the one hand and the health professional on the other hand will hopefully lead to a situation where the patient is able to look after his health (*d570*). The patient should be aware of the need to follow medical and other health advice and to avoid risks to health such as smoking and drinking alcohol. The possibility to influence and to improve the situation after therapy of a life-threatening disease such as HNC can generate motivation in different directions (*b130*) [48, 49].

Graphical modeling is one possibility to visualize associations of different areas of functioning in HNC patients and will help to get a deeper understanding for the rehabilitation of these patients. Our study could show different meaningful associations between areas of functioning in HNC patients after initial therapy. One central point could be visualized consisting of a circular path of *d330 Speaking*, *d350 Conversation*, *b510 Ingestion functions, s320 Structure of mouth*, and *b310 Voice functions*. Another important structure in the graph were the bow-shaped associations beginning with *d335 Producing nonverbal messages* to *b130 Energy and drive functions*. These findings can improve rehabilitation in the future and will be the basis for further studies.

Acknowledgements

We thank the members of our study centers for the patient recruitment and patient interviews. Furthermore, we want to thank Elisabeth Linseisen for the coordination of the different study centers.

Literaturverzeichnis

[1] A Kugaya, T Akechi, H Okamura, I Mikami, and Y Uchitomi. Correlates of depressed mood in ambulatory head and neck cancer patients. *Psychooncology*, 8:494–9, 1999.

[2] SN Rogers, G Humphris, D Lowe, JS Brown, and ED Vaughan. The impact of surgery for oral cancer on quality of life as measured by the Medical Outcomes Short Form 36. *Oral Oncol*, 34:171–9, 1998.

[3] DL Ronis, SA Duffy, KE Fowler, MJ Khan, and JE Terrell. Changes in quality of life over 1 year in patients with head and neck cancer. *Arch Otolaryngol Head Neck Surg*, 134:241–8, 2008.

[4] LQ Rogers, KS Courneya, KT Robbins, J Malone, A Seiz, L Koch, and K Rao. Physical activity correlates and barriers in head and neck cancer patients. *Support Care Cancer*, 16:19–27, 2008.

[5] Verdonck de Leeuw IM, Eerenstein SE, Van der Linden MH, Kuik DJ, de Bree R, and Leemans CR. Distress in spouses and patients after treatment for head and neck cancer. *Laryngoscope*, 117:238–41, 2007.

[6] G Stucki, T Ewert, and A Cieza. Value and application of the ICF in rehabilitation medicine. *Disabil Rehabil*, 24:932–8, 2002.

[7] WHO. International Classification of Functioning, Disability and Health: ICF, 2001.

[8] BR Pauloski, AW Rademaker, JA Logemann, and LA Colangelo. Speech and swallowing in irradiated and nonirradiated postsurgical oral cancer patients. *Otolaryngol Head Neck Surg*, 118:616–24, 1998.

[9] G Torrejano and I Guimaraes. Voice quality after supracricoid laryngectomy and total laryngectomy with insertion of voice prosthesis. *J Voice*, 23:240–6, 2009.

[10] C Callahan. Facial disfigurement and sense of self in head and neck cancer. *Soc Work Health Care*, 40:73–87, 2004.

[11] BA Murphy, JL Beaumont, J Isitt, AS Garden, CK Gwede, AM Trotti, RF Meredith, JB Epstein, QT Le, DM Brizel, LA Bellm, N Wells, and D Cella. Mucositis-related morbidity and resource utilization in head and neck cancer patients receiving radiation therapy with or without chemotherapy. *J Pain Symptom Manage*, 38:522–32, 2009.

[12] GW Blood, KC Simpson, M Dineen, SM Kauffman, and SC Raimondi. Spouses of individuals with laryngeal cancer: caregiver strain and burden. *J Commun Disord*, 27:19–35, 1994.

[13] MR Katz, N Kopek, J Waldron, GM Devins, and G Tomlinson. Screening for depression in head and neck cancer. *Psychooncology*, 13:269–80, 2004.

[14] R Strobl, G Stucki, E Grill, M Müller, and U Mansmann. Graphical models illustrated complex associations between variables describing human functioning. *J Clin Epidemiol*, (62):922–33, 2009.

[15] Kenneth A. Bollen. Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53(1):605–634, 2002.

[16] U Tschiesner, E Linseisen, S Baumann, V Siedek, K Stelter, A Berghaus, and A Cieza. Assessment of functioning in patients with head and neck cancer according to the International Classification of Functioning and Disability and and Health (ICF): a multicenter study. *Laryngoscope*, 119:915–23, 2009.

[17] U Tschiesner, S Rogers, A Dietz, B Yueh, and A Cieza. Development of ICF core sets for head and neck cancer. *Head Neck*, 32:210–20, 2010.

[18] WHO. The ICF checklist version 2.1a. Clinical form for International Classification of Functioning and Disability and Health, 2002.

[19] U Tschiesner, SN Rogers, U Harreus, A Berghaus, and A Cieza. Content comparison of quality of life questionnaires used in head and neck cancer based on the International Classification of Functioning and Disability and Health: a systematic review. *Eur Arch Otorhinolaryngol*, 265:627–37, 2008.

[20] S Becker, I Kirchberger, A Cieza, A Berghaus, U Harreus, O Reichel, and U Tschiesner.

[21] U Tschiesner, S Becker, and A Cieza. Health professional perspective on disability in head and neck cancer. *Arch Otolaryngol Head Neck Surg*, 136:576–83, 2010.

[22] DB Rubin and N. Schenker. Multiple imputation in health-care databases: an overview and some applications. *Stat Med*, 10:585–98, 1991.

[23] Joseph G. Ibrahim, Ming-Hui Chen, Stuart R. Lipsitz, and Amy H. Herring. Missing-Data Methods for Generalized Linear Models: A Comparative Review. *Journal of the American Statistical Association*, 100:332–346, 2005.

[24] N. J. Horton and K. P. Kleinman. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *Am Stat*, 61(1):79–90, 2007.

[25] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.

[26] S. van Buuren and C.G.M. Oudshoorn. Multivariate imputation by chained equations: Mice V1.0 Users Manual, 2000.

[27] R. Tibshirani. The lasso method for variable selection in the Cox model. *Stat Med*, 16(4):385–95, 1997.

[28] Jeff Gentry, Li Long, Robert Gentleman, and Seth Falcon. Rgraphviz: Provides plotting capabilities for R graph objects, 2008.

[29] Tomihisa Kamada and Satoru Kawai. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31:7–15, 1989.

[30] A. R. Donders, G. J. van der Heijden, T. Stijnen, and K. G. Moons. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*, 59(10):1087–91, 2006.

[31] F. M. Shrive, H. Stuart, H. Quan, and W. A. Ghali. Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. *BMC Med Res Methodol*, 6:57, 2006.

[32] G. J. van der Heijden, A. R. Donders, T. Stijnen, and K. G. Moons. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol*, 59(10):1102–9, 2006.

[33] J Bindewald, J Oeken, D Wollbrueck, C Wulke, A Dietz, E Herrmann, R Schwarz, and S Singer. Quality of life correlates after surgery for laryngeal carcinoma. *Laryngoscope*, 117:1770–6, 2007.

[34] Olthoff A, Steuer-Vogt MK, Licht K, Sauer-Goenen M, Werner C, and Ambrosch P. Quality of life after treatment for laryngeal carcinomas. *ORL J Otorhinolaryngol Relat Spec*, 68:253–8, 2006.

[35] Vilaseca I, Chen AY, and Backscheider AG. Long-term quality of life after total laryngectomy. *Head Neck*, 28:313–10, 2006.

[36] RK Wong, ES Poon, CY Woo, SC Chan, ES Wong, and AW Chu. Speech outcomes in Cantonese patients after glossectomy. *Head Neck*, 29:758–64, 2007.

[37] KJ Lorenz, K Groll, AH Ackerstaff, FJ Hilgers, and H Maier. Hands-free speech after surgical voice rehabilitation with a Provox voice prosthesis: experience with the Provox FreeHands HME tracheostoma valve system. *Eur Arch Otorhinolaryngol*, 264:151–7, 2007.

[38] MB Happ, TK Roesch, and SH Kagan. Patient communication following head and neck cancer surgery: a pilot study using electronic speech-generating devices. *Oncol Nurs Forum*, 32:1179–87, 2005.

[39] MA List and SP Bilir. Evaluations of quality of life and organ function. *Semin Oncol*, 31:827–35, 2004.

[40] M Roing, JM Hirsch, and I Holmstrom. Living in a state of suspensionda phenomenological approach to the spouse's experience of oral cancer. *Scand J Caring Sci*, 22:40–7, 2008.

[41] MJ Ramirez, EE Ferriol, FG Domenech, MC Llatas, MM Suarez-Varela, and RL Martinez. Psychosocial adjustment in patients surgically treated for laryngeal cancer. *Otolaryngol Head Neck Surg*, 129:92–7, 2003.

[42] RO Seidl and R Nusser-Muller-Busch. Posttreatment rehabilitation of swallowing in patients with head and neck cancer. *Laryngorhinootologie*, 86:846–52, 2007.

[43] P Boscolo-Rizzo, M Stellin, R Fuson, C Marchiori, A Gava, and MC Da Mosto. Long-term quality of life after treatment for locally advanced oropharyngeal carcinoma: surgery and postoperative radiotherapy versus concurrent chemoradiation. *Oral Oncol*, 45:953–7, 2009.

[44] B Scott, C Butterworth, D Lowe, and SN Rogers. Factors associated with restricted mouth opening and its relationship to health-related quality of life in patients attending a Maxillofacial Oncology clinic. *Oral Oncol*, 44:430–8, 2008.

[45] JC de Vicente, L de Villalain, A Torre, and I Pena. Microvascular free tissue transfer for tongue reconstruction after hemiglossectomy: a functional assessment of radial forearm versus anterolateral thigh flap. *J Oral Maxillofac Surg*, 66:2270–5, 2008.

[46] P Delaere, A Goeleven, VV Poorten, R Hermans, R Hierner, and J Vranckx. Organ preservation surgery for advanced unilateral glottic and subglottic cancer. *Laryngoscope*, 117:1764–9, 2007.

[47] PS Mehta and LBJ Harrison. Function and organ preservation in adult cancers of the head and neck. *Expert Rev Anticancer Ther*, 7:361–71, 2007.

[48] S Singer, M Fuchs, A Dietz, E Klemm, U Kienast, A Meyer, J Oeken, R Täschner, C Wulke, and R Schwarz. Relevance of psychosocial factors in speech rehabilitation after laryngectomy. *Laryngorhinootologie*, 86:867–74, 2007.

[49] J Zimmer-Nowicka, E Narzekalak, D Kaczmarczyk, I Niedzwieckac, and A Morawiec-Bajda. The role of a psychologist in a rehabilitation of the patients with laryngeal cancer after total laryngectomy with implantation of the voice prostheses. *Otolaryngol Pol*, 60:343–7, 2006.

DANKSAGUNG

Mein erster Dank gilt meiner Betreuerin Eva Grill für die tolle und kompetente Betreuung. Ohne ihre Unterstützung, Aufmunterung und ihren Glauben an mich, wäre diese Arbeit nicht möglich gewesen. Alles wird gut? Alles ist gut.

Desweiteren möchte ich mich herzlich bedanken...

- bei Prof. Dr. Ulrich Mansmann, der mit seinen Ideen diese Arbeit erst ins Rollen gebracht hat
- bei Prof. Dr. Gerold Stucki, dessen Begeisterungsfähigkeit der Arbeit Leben eingehaucht hat
- bei Dr. rer. medic. Martin Müller, Schwester und fast MPH, für den täglichen fruchtbaren Meinungsaustausch
- bei Rene Mittrach für die hervorragende Betreuung des Damenvolleyballteams (und mehr)
- bei Emmi Beck und Kaitlen Howell for improving my written and spoken english
- bei Sven Becker und Jan Reinhardt für die gute und konstruktive Zusammenarbeit
- bei allen anderen lieben Kolleginnen und Kollegen die mich inspiriert und unterstützt haben.

Mein ganz besonderer Dank gilt natürlich meinen Eltern, die mir diese Ausbildung erst ermöglicht haben und meiner Familie Anneke, Nora und Moritz für ihre Liebe, Aufmunterung und Ablenkung!