

Untersuchung der Retestrelibilität des Impact of Vision Impairment-Fragebogens im zeitlichen Verlauf und vergleichend zwischen drei Befragungsmodi

Dissertation

zur Erlangung des Doktorgrades (Dr. med.)

der Medizinischen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität

Bonn

Reglind Andrea Dorothee Ost

aus Oldenburg

2026

Angefertigt mit der Genehmigung
der Medizinischen Fakultät der Universität Bonn

1. Gutachter: Prof. Dr. Dr. Robert P. Finger
2. Gutachter: Prof. Dr. Matthias Schmid

Tag der Mündlichen Prüfung: 18.03.2026

Aus der Klinik und Poliklinik für Augenheilkunde

Inhaltsverzeichnis

Abkürzungsverzeichnis	5
1. Deutsche Zusammenfassung	6
1.1 Einleitung	6
1.2 Material und Methoden	9
1.2.1 Rekrutierung	9
1.2.2 Entwicklung und Aufbau des Fragebogens	10
1.2.3 Studienablauf	11
1.2.4 Datenauswertung	12
1.2.4.1 Rasch-Analyse mittels Winsteps	12
1.2.4.2 Statistische Analysen	14
1.2.4.3 Analysen zur Darstellung systematischer Verzerrungen	16
1.3 Ergebnisse	16
1.3.1 Merkmale der Studienkohorte	16
1.3.2 Ablauf der Befragung	17
1.3.3 Psychometrische Auswertung des IVI	18
1.3.4 Reliabilitätsanalyse	19

1.3.4.1	Retest	20
1.3.4.2	Vergleich der Befragungsmodi	21
1.3.4.3	Störfaktoren-Analyse	22
1.4	Diskussion	22
1.5	Zusammenfassung	26
1.6	Literaturverzeichnis der deutschen Zusammenfassung	28
2.	Veröffentlichungen	33
3.	Erklärung zum Eigenanteil	34
4.	Danksagung	35
5.	Publikationen (PDF-Version)	36

Abkürzungsverzeichnis

CoR	Wiederholbarkeitskoeffizient (<i>Coefficient of Repeatability</i>)
DIF	Gruppenabhängiger Indikator (<i>Differential Item Functioning</i>)
ICC	Intraklassenkorrelationskoeffizient (<i>Intra-class Correlation Coefficient</i>)
IVI	Impact of Vision Impairment-Fragebogen
IVI_E	Emotionale Subskala des Impact of Vision Impairment Fragebogens
IVI_F	Funktionale Subskala des Impact of Vision Impairment Fragebogens
MNSQ	Mittelwert-Quadrat-Standardresiduen (<i>Mean square standardized residuals</i>)
PCA	Hauptkomponentenanalyse (<i>Principal Component Analysis</i>)
PR	Personenreliabilitätskoeffizient (<i>Person Reliability</i>)
PROMs	Patientenberichtete Endpunkte (<i>Patient-Reported Outcome Measures</i>)
PSI	Personentrennungsindex (<i>Person Separation Index</i>)
VRQoL	Sehbezogene Lebensqualität (<i>Vision-Related Quality of Life</i>)

1. Deutsche Zusammenfassung

1.1 Einleitung

Essentielle Grundlagen des medizinischen Handelns sind die Entscheidungen über die diagnostischen Verfahren und die Auswahl des Therapiekonzeptes. Traditionell erfolgte insbesondere die Therapieentscheidung entsprechend dem paternalistischen Prinzip durch das ärztliche Personal (Joosten et al., 2008). Im Gegensatz dazu wird in der Medizin der letzten Jahre zunehmend das Modell der partizipativen Entscheidungsfindung genutzt, sodass sich auch in der medizinischen Publikationsdatenbank PubMed innerhalb der letzten fünf Jahre über 12 000 Publikationen zum Thema „shared decision making“ (Stand 12.03.2024) finden (Bouaziz et al., 2022). Das Ziel ist die gemeinsame Entscheidung von Arzt/Ärztin und Patient/in für ein Therapiekonzept, wobei die Aufklärung des/der Patienten/-innen im Mittelpunkt steht (Joosten et al., 2008; Spatz et al., 2017). Die vermehrte Einbeziehung der Patientensicht und damit auch der Lebensqualität führt, insbesondere bei Patienten/-innen mit chronischen Erkrankungen, zu einer erhöhten Patientenzufriedenheit (Shay und Lafata, 2015; Tapp et al., 2017; Xu et al., 2017; Joosten et al., 2008).

Die WHO-Definition beschreibt Lebensqualität als „die Wahrnehmung der Lebenssituation des/der Einzelnen im Kontext der Kultur und Wertesysteme, in denen er/sie lebt, und in Bezug auf seine/ihre Ziele, Erwartungen, Normvorstellungen und persönlichen Anliegen“ (World Health Organization, 1998). Patientenberichtete Endpunkte (PROMs, englisch patient-reported outcome measures) ermöglichen eine Erfassung dieser Bereiche und haben sich vor diesem Hintergrund insbesondere in der Arzneimittelregulatorik und Forschung, aber auch in der Patientenversorgung etabliert (Braithwaite et al., 2019; Dean et al., 2017; Loukanova und Bridges, 2008). Bei PROMs handelt es sich um standardisierte Fragebögen, welche zum Beispiel den Gesundheitszustand aus Patientensicht erfassen und beurteilen lassen (Braithwaite et al., 2019). Im Bereich der Augenheilkunde werden insbesondere Fragen zur Alltagsrelevanz von Seheinschränkungen, Symptomen und dem Aspekt der sehbezogenen Lebensqualität (VRQoL, englisch Vision-related Quality of Life) eingesetzt (Black, 2013; Denniston et al., 2014). Dem Sehen wird von vielen Menschen ein sehr hoher Stellenwert beigemessen

und bereits leichte Seheinschränkungen können zu einer deutlich verschlechterten sehbezogenen Lebensqualität führen, weshalb die VRQoL als nützliches Instrument zur Gewinnung zusätzlicher Informationen bei der Prüfung der Sehfunktion gesehen wird (Finger et al., 2011; Knauer und Pfeiffer, 2008; Varma et al., 2006). Die konventionellen Funktionstestungen wie beispielsweise die bestkorrigierte Sehschärfe oder die Gesichtsfeldprüfung gehören damit weiterhin zur ophthalmologischen Hauptdiagnostik, reichen aber nicht aus, um Einzelheiten über die individuelle Krankheitssituation zu erfahren und letztendlich die bestmögliche Therapie im Sinne des/der Patienten/innen anzubieten (Dean et al., 2017; Denniston et al., 2014; Massof und Rubin, 2001). In der Ophthalmologie sind PROMs wie der Impact of Vision Impairment (IVI) Fragebogen oder der National Eye Institute Visual Function Questionnaire-25 (NEI-VFQ-25) gängige Instrumente zur Beurteilung der Lebensqualität in Bezug auf die Sehfunktion (Braithwaite et al., 2019; Boer et al., 2004; Luu et al., 2020). Analog zu psychophysikalischen Funktionstests unterliegen PROM-Instrumente einem sorgfältigen Validierungsprozess, wobei man zwischen qualitativen und quantitativen Validierungsschritten unterscheidet. Qualitative Schritte dienen der inhaltlichen Validierung (Erheben die Fragen den für die Lebensqualität relevanten Inhalt?), während quantitative Validierungsschritte die statistische Messgüte eines PROM überprüfen (Besteht eine ausreichende Reliabilität? Ist das Instrument in sich konsistent? Welche Ergebnisse liefert der Fragebogen im Vergleich zu anderen Beurteilungsverfahren?) (Frost et al., 2007; Lamoureux et al., 2007b; Massof und Rubin, 2001).

In unserer Studie untersuchten wir den Impact of Vision Impairment-Fragebogen (IVI), welcher inzwischen in Rehabilitations- und Therapiestudien weit verbreitet eingesetzt wird und eines der drei am häufigsten verwendeten VRQoL-Instrumente ist (Boer et al., 2004; Finger et al., 2014; Hassell et al., 2006; Kanellopoulos, 2019; Lamoureux et al., 2007b; Luu et al., 2020; Pondorfer et al., 2019; Tan et al., 2019). Die Entwicklung des IVI erfolgte primär mit dem Ziel festzustellen, ob Patienten/-innen aufgrund einer verminderten Sehfähigkeit in ihrer Alltagsbewältigung eingeschränkt sind und Rehabilitationsmaßnahmen benötigen (Hassell et al., 2000; Keeffe et al., 1999). Seit der Einführung wurde das Grundgerüst mehrfach vor allem durch den Einsatz moderner psychometrischer Messverfahren zu dem aktuell gebräuchlichen IVI mit 28 Items modifiziert (Lamoureux et al., 2006; Lamoureux et al., 2007c). Bislang wurden in drei

Studien, die mit insgesamt 102 Retest-Teilnehmern/-innen (n= 22; 60; 20) vergleichsweise kleine Gruppen darstellten, Daten zur Retestreliabilität des IVI gewonnen (Marakis et al., 2020; Ratanasukon et al., 2016; Weih et al., 2002). Angesichts dieser limitierten Evidenz bestand die Notwendigkeit, diese erneut in einer größeren Kohorte zu überprüfen. Neben der Wiederholbarkeit über verschiedene Zeitpunkte hinweg, stellt auch die Konsistenz der Ergebnisse hinsichtlich verschiedener Befragungsmodi einen weiteren bedeutenden Aspekt von Fragebogenstudien dar. Daten zur Untersuchung des möglichen Vorhandenseins einer systematischen Verzerrung durch die jeweilige Wahl des Befragungsmodus beschränken sich auf eine Studie mit 31 Teilnehmern/-innen hinsichtlich des Vergleichs zwischen Papierfragebogen und einer Interviewbefragung (Weih et al., 2002). Da bezüglich einer Interviewer/in-geführten telefonischen Befragung und der Selbstbeurteilung eine bekannte Tendenz zur Beantwortung der Fragen im Sinne einer besseren Lebensqualität beim Interviewmodus bekannt ist, sahen wir auch hier Bedarf für eine weitere Studie (Frost et al., 2001; Hanmer et al., 2007; Wolffsohn et al., 2000). Aufgrund zahlreicher Vorteile wie beschleunigte und insgesamt höhere Rücklaufquoten, Kostensenkung und Reduktion von Arbeitsaufwand bei der Datenverarbeitung erfreuen sich insbesondere elektronische Befragungsmethoden deutlich zunehmender Beliebtheit. In Anbetracht der Bedürfnisse sehr unterschiedlicher Patientengruppen (Alter, Sehvermögen, Mobilität, elektronische Kenntnisse u.v.m.) erscheint es sinnvoll, die Fragebögen in verschiedenen Administrationsmodi, wie beispielsweise im Interviewmodus per Face-to-Face-Kontakt oder per Telefon, als selbst-administrierten Fragebogen in Papierform oder als elektronische Version zur Verfügung stellen zu können (Evans und Mathur, 2005).

Ein systematischer Vergleich des IVI zwischen elektronischen, papierbasierten und Interview-basierten Methoden wurde in keiner der oben genannten Studien durchgeführt, dabei könnte eine diesbezügliche Auswertung zu enormen Erleichterungen führen. Trotz der Vielfalt der Anwendungsmethoden könnte unter anderem mit Hilfe dieser Studie in Zukunft die Möglichkeit eröffnet werden, Studien mit unterschiedlichen Administrationsmodi adäquat zu vergleichen und damit kumulativ einen höheren Erkenntnisgewinn aus jeder Studie zu erzielen.

Da in Therapie- oder Langzeitstudien Veränderungen der VRQoL zuverlässig im Messinstrument registriert werden sollten und nicht durch Störfaktoren hervorgerufen

oder maskiert werden dürfen (Vet et al., 2001), wurden in dieser Studie die zwei folgenden Fragestellungen bearbeitet:

- 1) Liegt beim IVI auch in einer Kohorte mit adäquatem Stichprobenumfang eine gute Retestreliaibilität vor?
- 2) Liefern verschiedene Befragungsmodi dieses PROMs vergleichbare Ergebnisse?

Zur weiteren wissenschaftlichen Bearbeitung und Überprüfung wurden die Fragestellungen in die folgenden Hypothesen umformuliert:

Hypothese 1: Es bestehen keine Unterschiede zwischen den Antworten der Erst- und der Folgebefragung des IVI.

Hypothese 2: Es bestehen keine Unterschiede zwischen den Antworten bei der Befragung mittels unterschiedlicher Administrationsmodi (Telefonische Version, Papierversion, Elektronische Version).

Im Folgenden wird aus Gründen der besseren Lesbarkeit das generische Maskulinum verwendet. Damit wird auf die gleichzeitige Verwendung der Sprachformen männlich, weiblich und divers verzichtet, wobei sich die Personenbezeichnungen auf alle Geschlechter beziehen.

1.2 Material und Methoden

1.2.1 Rekrutierung

Vor Beginn unserer Studie an der Universitäts-Augenklinik in Bonn erfolgte die Antragsstellung zur Durchführung der prospektiven, klinischen Studie bei der Ethikkommission. Unter der Referenznummer 130/16 wurde der Studie die Zustimmung durch die Ethikkommission erteilt. Das Studienprotokoll entspricht den Grundsätzen der Deklaration von Helsinki. Es folgte die Rekrutierungsphase. Hierzu wurden Patienten, die im letzten halben Jahr mindestens einmal in der Augenklinik einen Termin wahrgenommen haben, über das klinikinterne Patientensystem telefonisch oder direkt im Rahmen der ambulanten (Kontroll-) termine zur Studienteilnahme angefragt. Um in die Studie eingeschlossen zu werden, waren die folgenden Kriterien zu erfüllen:

- schriftliches Einverständnis zur Studienteilnahme
- ≥ 18 Jahre
- ausreichenden Deutschkenntnissen
- stabile, chronische Augenerkrankung(en)
- mindestens zwei vorherige Untersuchungen in der UKB-Augenklinik
- letzte ophthalmologische Untersuchung maximal sechs Monate vor Studieneintritt

Ausgeschlossen wurden Patienten bei folgenden Situationen:

- akut aufgetretene Augenerkrankung in den sechs Wochen vor/während der Studienteilnahme
- Augenoperationen oder Eingriffe in den sechs Wochen vor/während der Studienteilnahme (Ausnahme: intravitreale Injektionen)

Die Patientenrekrutierung und Teilnehmerbefragung erfolgte in einem zeitlichen Rahmen von April bis Dezember 2020.

1.2.2 Entwicklung und Aufbau des Fragebogens

Unter der Vielzahl an Fragebögen, die in den letzten Jahrzehnten für die visusbezogene Lebensqualität entwickelt wurden, ist der IVI durch die Berücksichtigung von Einschränkungen bei der gesellschaftlichen Teilhabe und damit der Anwendbarkeit im Rehabilitationskontext klinisch besonders relevant (Weih et al., 2002). Der „Impact of Vision Impairment“- Fragebogen umfasst sogenannte Patientenberichtete Endpunkte (PROMs) und misst damit aus Patientenperspektive verschiedene Bereiche der sehbezogenen Lebensqualität (Lamoureux et al., 2006). Es fließen insbesondere auch emotionale Aspekte ein, welche für die Rehabilitation bei ophthalmologischen Erkrankungen eine große Rolle spielen (Hassell et al., 2000; Lamoureux et al., 2006). Im Jahr 1999 wurde die erste Version mit zunächst 72 Items zur Befragung von Patienten im Interviewmodus erstellt (Keeffe et al., 1999). Aufgrund von Validierungsstudien sowie der Praktikabilität bei der Fragebogenadministration erfolgte in den folgenden Jahren eine Kürzung der Fragen auf zunächst 32 und dann 28 Items (Hassell et al., 2000; Lamoureux et al., 2006; Weih et al., 2002). Auch die Anzahl der Antwortmöglichkeiten wurde mithilfe psychometrischer Analysen auf drei bis vier aktive Antwortmöglichkeiten umstrukturiert

(Lamoureux et al., 2006). Unserer Studie legten wir die validierte deutsche Version mit 28 Items zugrunde (Finger et al., 2011). Diese beinhaltet 28 Fragen mit jeweils fünf Antwortoptionen für die Fragen 1-13 und vier Antwortoptionen für die Fragen 14-28. Es ist zwischen den Antwortoptionen „überhaupt nicht“, „ab und zu“, „häufig“, „sehr häufig“ und bei den Fragen 1-13 zusätzlich der fünften Option „Ich mache das aus anderen Gründen nicht“ zu wählen (Finger et al., 2011). Diese zusätzliche Antwortoption wurde in unserer Auswertung aus Übersichtsgründen als fehlend betrachtet. Bei der Originalversion des IVI fand sich eine dreigeteilte Fragebogenstruktur in „Reading and accessing information“, „Mobility and independence“ and „Emotional well-being“. In der deutschen Validierungsstudie zeigte sich jedoch eine Zweiteilung, welche wir in unseren Auswertungen ebenfalls übernommen haben. Damit ergab sich für die ersten zwanzig Fragen die Zuteilung in die Untergruppe praktischer Fähigkeiten und Beeinträchtigungen = „Funktionaler IVI“ und für die Fragen 21-28 die Zuteilung zum psychologisch-emotionalen Bereich = „Emotionaler IVI“. Die Beantwortung der Fragen erfolgte sowohl in der Ursprungsversion von Keeffe et al. sowie auch in der deutschen Validierungsstudie von Finger et al. im Interview-Modus (Finger et al., 2011; Keeffe et al., 1999).

1.2.3 Studienablauf

Der IVI wurde den Patienten im Studienverlauf wiederholt in bis zu drei unterschiedlichen Befragungsmodi dargereicht. Weder dem Patienten noch dem Interviewer lagen bei einer erneuten Befragung die zuvor gegebenen Antworten vor. Das Studienprotokoll sah zwei Telefoninterviews, eine postalische Befragung mit ausgedrucktem Fragebogen sowie eine Online-Befragung mit elektronischem Fragebogen vor. Bei vorzeitiger Beendigung der Studienteilnahme nach weniger als zwei Befragungen erfolgte der Studienausschluss. Das maximale Antwortintervall zwischen den jeweiligen Befragungen wurde auf zehn Wochen festgelegt und Überschreitungen dieses Intervalls führten ebenfalls zum Studienausschluss. Die Entscheidung über den ersten Befragungsmodus („Papier“ oder „Telefon“), wurde durch den Probanden gefällt.

Im folgenden Abschnitt erfolgt eine detaillierte Beschreibung der jeweiligen Befragungsmodi:

„Papier“: Der Fragebogen wurde dem Probanden postalisch zugesandt und handschriftlich ausgefüllt an uns zurückgesendet. Diese Antworten wurden händisch, unter zweimaliger Kontrolle, in eine Studiendatenbank übertragen.

„Telefon“: Es erfolgte ein ca. 30-minütiges Telefoninterview beginnend mit der Nennung der vorgegebenen Antwortmöglichkeiten. Es folgte die Fragestellung und anschließend die erneute Nennung der Antwortmöglichkeiten. Der Proband wählte eine der Antwortmöglichkeiten. Diese wurde durch den Interviewer in einer Studiendatenbank vermerkt. Die zweite Befragung mittels eines Telefoninterviews erfolgte nach dem gleichen Schema.

„Elektronisch“: Es wurde ein Link per Mail verschickt und bei Anklicken des Links öffnete sich ein Online-Fragebogen, bei welchem die Probanden durch Anklicken die Antworten eigenständig auswählten. Nach Beendigung des Fragebogens war dieser für die Auswertungen zugänglich.

Die weiteren auswertungsrelevanten Patientendaten wurden als strukturierte Fragebögen als Papierformular mit versendet und bei fehlenden Eintragungen telefonisch abgefragt. Hierbei lag der Fokus insbesondere auf den soziodemografischen Daten und der allgemeinen Krankengeschichte. Die ophthalmologische Vorgeschichte inklusive Diagnosen und Visusangaben der letzten Kontrolluntersuchung wurde aus den Patientenakten der Uniklinik Bonn entnommen.

1.2.4 Datenauswertung

1.2.4.1 Rasch-Analyse mittels Winsteps

Die Antworten des Fragebogens lagen als Ordinaldatensatz vor und waren somit einer parametrischen statistischen Analyse nicht zugänglich. Um dieser Problematik zu entgegen erfolgte eine Umwandlung der vorliegenden Daten mittels Rasch-Analyse über die Software Winsteps (Winsteps Software, Version 3.92.1.2; Winsteps, Chicago, IL) unter Verwendung der Andrich-Rating Skala (Linacre, 2006; Wright und Linacre, 1989).

Unter dem Rasch-Modell versteht man ein psychometrisches Verfahren zur Analyse, wie gut ein Messwert zu dem zugrundeliegenden Merkmal passt. Es schätzt die Schwierigkeit der Frage (Itemmesswert) im Verhältnis zur Fähigkeit der Person (Personenmesswert),

indem beide auf derselben linearen Skala platziert werden (Wright und Linacre, 1989). Die Daten werden hierdurch von ordinalen Likert-Skalen in ein Pseudo-Intervallskalenniveau gebracht und die Personenmesswerte sind dementsprechend von negativen zu positiven Werten geordnet (Wright und Linacre, 1989). Durch Umkehrung der Werte schafften wir mit einer Zuordnung von niedrigeren Werten zu einer schlechteren Sehfähigkeit eine intuitivere Interpretationsbasis. Zudem wurde ein Ankerverfahren angewendet, wobei die Antworten des ersten Telefoninterviews als Grundlage der Analyse dienten, um eine direkte Vergleichbarkeit der Ergebnisse zu erzielen. Es wurde wiederholt dargelegt, dass ein Face-to-Face-Interview und ein telefonisches Interview sehr vergleichbare Ergebnisse liefern (Herzog und Rodgers, 1988; Hochstim, 1967). Das Telefoninterview wurde daher als bestmögliche Baseline gewählt, da die zugrunde gelegte Validierungsstudie ebenfalls im Interview-Modus erfolgte (Finger et al., 2011).

Zur Durchführung und Interpretation der Rasch-Analyse berücksichtigten wir verschiedene Parameter. Die Untersuchung der Anordnung der Schwellenwerte (threshold ordering) ist notwendig, um zu entscheiden, ob die Itemanordnung zu der Kategoriestructur passt (Lamoureux et al., 2006). Unpassende Werte werden aufgrund von Über- oder Unterbelegung oder Diskrimierungsschwierigkeiten innerhalb der Antwortkategorien offensichtlich. Falls dies bei Durchsicht der Kategorien auffällig ist, sollten diese zusammengelegt werden, um das Modell zu verbessern (Lamoureux et al., 2006). Besondere Bedeutung wird dem Personentrennungsindex (PSI) zugeschrieben, denn um eine gute Differenzierung der Personen in mindestens zwei verschiedene Gruppen zu erreichen, wird ein Wert von über 2,0 (1,5 bis 2,0 akzeptabel) und zudem ein Personenreliabilitätskoeffizient (PR) $\geq 0,8$ (0,6 bis 0,8 akzeptabel) angestrebt (Bond und Fox, 2001; Boone und Noltemeyer, 2017; Wright und Masters, 1982). Die Bestimmung der Eindimensionalität ist ein weiterer bedeutsamer Aspekt für die Analyse, denn diese zeigt an, dass nur ein einziges zugrundeliegendes Merkmal gemessen wird und dass die Items zu diesem Merkmal passen. Die Einordnung diesbezüglich erfolgt mittels der Infit- und Outfit-Mittelwert-Quadrat-Standardresiduen-Statistik (englisch mean square standardized residuals, MNSQ). Infit-Misfit ist interessant bei Antworten, die nahe dem erwarteten Personenmaß liegen. Outfit-Misfit ist relevant für Verhalten, das weit von dem erwarteten normalen Personenmaß entfernt ist. Der Infit- oder Outfit-MNSQ-Wert von 1 gilt als ideal, aber bei einer Umfrage mit Ratingskala gelten Werte bis zu 1,4 als

angemessen (Boone und Noltemeyer, 2017; Wright und Linacre, 1994). Höhere Werte würden aufgrund eines ausgeprägten Hintergrundrauschens das Messsystem in seiner Zuverlässigkeit beeinträchtigen, und Werte unter 0,6 werden als überangepasst betrachtet, wodurch die weitere statistische Analyse jedoch nicht beeinträchtigt werde (Wright und Linacre, 1994). Die Prüfung auf lokale Unabhängigkeit erfolgt mit der Hauptkomponentenanalyse (PCA). Die PCA der Residuen für den ersten Faktor sollte über 50 % liegen und der 1. Kontrast der Residuen sollte kleiner als 2,0 Eigenwert sein, da sonst der Ausschluss einer weiteren interagierenden Dimension erfolgen sollte (Linacre, 2006). Bei dem Verdacht auf eine weitere interagierende Dimension ist eine Betrachtung der Tabellen auf der Suche nach möglichen Clustern wichtig. Eine visuelle Darstellung der Differenz zwischen Personen- und Item-Mittelwert-Logits erfolgt über die Wright-Karte. Sie vermittelt einen Eindruck davon, ob die Personen-Item-Zielsetzung geeignet ist. Bei Werten nahe null kann von einer exakt passenden Ausrichtung ausgegangen werden und daher davon, dass die sehbedingten Alltagseinschränkungen der Personen zu den Schwellenwerten passen und die Items gut verteilt sind. Im Gegensatz dazu weisen Werte $>1,0$ Logits auf eine fehlerhafte Einteilung hin mit einer Mehrheit der Personen, die die Items besser oder schlechter beantwortet als aufgrund der Verteilung der Items zu erwarten wäre. Zur Berücksichtigung von Unterschieden zwischen Geschlechts- und Altersgruppen erfolgte eine differentielle Itemfunktionsanalyse (DIF). Als auffällig betrachtet man Werte von über 0,64, da diese zu der Annahme führen, dass eine Untergruppe bei ansonsten gleichen sehbezogenen Alltagsfähigkeiten aufgrund des Parameters „Alter“ oder „Geschlecht“ systematisch anders auf ein bestimmtes Item reagiert. Die praktische Relevanz im Sinne einer signifikanten Abweichung wurde ermittelt, indem ein Vergleich der Personenmesswerte sowohl mit und ohne Beibehaltung der abweichenden Werte anhand der Intraklassenkorrelationskoeffizienten (ICC) durchgeführt wurde.

1.2.4.2 Statistische Analysen

Zur Klärung des Studienziels mit Aufdeckung möglicher Abweichungen zwischen den wiederholten telefonischen Befragungen sowie zur Identifikation von Abweichungen bei der Fragebogenadministration mittels verschiedener Methoden erfolgte eine statistische Auswertung des - nach Rasch-Analyse - intervallskalierten Datensatzes. Die

Durchführung der statistischen Analysen erfolgte mit SPSS (Version 27 Software, IBM SPSS Inc., Armonk, NY) und R (Version 4.1.2), wobei eine statistische Signifikanz bei p-Werten $<0,05$ angenommen wurde. Mehrfach wurde aufgrund von multiplem Testen eine Bonferroni-Korrektur bei der Signifikanztestung angewandt.

- a) Intraklassenkorrelationskoeffizient (englisch intra-class correlation coefficient, ICC): Der ICC ist das Maß eines parametrischen Verfahrens zur Reliabilitätstestung. Bei der Intraklassenkorrelation wird der Grad der Übereinstimmung zwischen zwei oder mehr Messungen festgestellt. Es liegen mehrere Möglichkeiten der Berechnung des ICC anhand von Mittelwertquadraten vor. Bekannt sind dabei die ursprüngliche Einteilung nach Shrout und Fleiss in sechs Unterkategorien und dann die Erweiterung von McGraw and Wong in zehn verschiedene Berechnungsgruppen (McGraw und Wong, 1996; Shrout und Fleiss, 1979). Hierbei wählten wir das Format „Two-way mixed effects, absolute agreement“, da wir eine wiederholte Testung (= two way mixed effects) durchführten und uns das Ausmaß, in dem die zu vergleichenden Werte übereinstimmten, interessierte (= absolute agreement) (Koo und Li, 2016; McGraw und Wong, 1996). Die Auswertung des ICC erfolgte anhand der Einordnung in die vier Kategorien „schlecht“ ($<0,40$), „mittelmäßig“ ($0,40 \leq \text{ICC} \leq 0,59$), „gut“ ($0,60 \leq \text{ICC} \leq 0,74$) und „ausgezeichnet“ ($0,75 \leq \text{ICC} \leq 1,00$) (Cicchetti und Sparrow, 1981). Zur Festlegung des Konfidenzintervalls wurde als Mittelwert der Unterschied zwischen den Messungen \pm Standardabweichung $\times 1,96$ berechnet (Giavarina, 2015).
- b) Wiederholbarkeitskoeffizient (englisch coefficient of repeatability, CoR): Hierbei handelt es sich um einen Wert zur Darstellung der absoluten Zuverlässigkeit, welcher sich besonders zum Vergleich mit anderen Studien anbietet. Dafür erfolgt die Subtraktion der oberen Mittelwerte der Differenzen, wobei der ermittelte Wert 1,96 Standardabweichungen entspricht (Bland und Altman, 1986; Vaz et al., 2013). Zudem ermöglicht die Berechnung dieses Wertes eine visuelle Darstellung der Ergebnisse mittels Bland-Altman-Plots.
- c) Regressionsanalyse nach Deming: Das lineare Regressionsmodell wurde zum Vergleich der Personenmesswerte zwischen verschiedenen Befragungszeitpunkten beziehungsweise -modi angewandt. Es berücksichtigt die

Messfehler in der abhängigen und unabhängigen Variable, wobei zur Interpretation der y-Achsenabschnitt auf signifikante Abweichungen von 0 (englisch intercept) und die Steigung auf signifikante Abweichungen von 1 (englisch slope) geprüft wurden (Ciccione und Dehaene, 2021; Deming, 1943).

1.2.4.3 Analysen zur Darstellung systematischer Verzerrungen

Der primäre Befragungsmodus („Telefon“ oder „Papier“) wurde von den Patienten selbst gewählt, sodass zum Ausschluss eines systematischen Bias ein Vergleich der „Telefongruppe“ mit der „Papiergruppe“ unter Verwendung des t-Testes und Chi-Quadrat-Testes erfolgte. Darüber hinaus nutzten wir den Mann-Whitney-U-Test, um die Relevanz mehrerer Einflussgrößen auf die mittleren Differenzen beim Retest-Vergleich und die mittleren Differenzen beim intermodalen Vergleich (Telefon- und Papiermodus) herauszuarbeiten. Die Analyse wurde unter Gebrauch der für Mehrfachtestungen korrigierenden Bonferroni-Methode durchgeführt. Als zu untersuchende, potentielle Störfaktoren wurden insbesondere die Sehschärfe und selbstregistrierte Hörprobleme berücksichtigt, um sicherzustellen, dass die Teilnehmer in der Lage waren, die gestellten Fragen angemessen wahrzunehmen und das Hören zudem auch, um eine möglicherweise veränderte Beantwortung der Fragen durch die Beeinträchtigung eines weiteren Sinnesorgans mit einzubeziehen. Zusätzlich wurden das Befragungsintervall sowie der primäre Befragungsmodus als potenzielle Störfaktoren eingeschätzt und in die Analyse eingeschlossen. Des Weiteren wurden auch selbst angegebene psychische Erkrankungen in die Analyse einbezogen, um potenzielle Auswirkungen von ausgeprägten emotionalen Belastungen oder anderen mentalen Beeinträchtigungen auf die Befragungsergebnisse zu erfassen.

1.3 Ergebnisse

1.3.1 Merkmale der Studienkohorte

Es wurden zunächst zufällig 416 erwachsene Patienten mit chronischen Augenerkrankungen identifiziert. Hiervon konnten 400 telefonisch erreicht werden und 246 verfügten über ausreichende Deutschkenntnisse und waren bereit an der Studie teilzunehmen. Davon wurden - nach Erfüllung der Mindestanforderung einer zweimaligen

Beantwortung des IVI - 216 Teilnehmern mit einem Durchschnittsalter von 67 ± 12 Jahren in die Analyse einbezogen (s. Tab. 1 (Terheyden et al., 2024)). Von den 30 ausgeschlossenen Teilnehmern waren 22 Probanden aus persönlichen Gründen nicht mehr zu einer Studienteilnahme bereit, 7 Probanden waren nicht mehr telefonisch erreichbar (mindestens 5 Anrufversuche) und ein Proband schied aufgrund einer Augenoperation im Studienintervall aus. Fragebögen mit einer unter 50-prozentigen Antwortrate wurden in der Studie nicht berücksichtigt ($n=5$). Die als Einschlusskriterium unserer Studienkohorte festgelegte chronische Augenerkrankung umfasste vitreoretinale Erkrankung, Glaukom, Katarakt oder ein weiteres selteneres ophthalmologisches Krankheitsbild (s. Supp. Tab.1 (Terheyden et al., 2024)). Insgesamt nahmen knapp 20 % mehr Frauen an der Studie teil und ein Viertel der Teilnehmer absolvierte mindestens die allgemeine Hochschulreife (s. Tab. 1 (Terheyden et al., 2024)).

1.3.2 Ablauf der Befragung

Die erste Befragung erfolgte je nach Patientenwunsch bei 98 Patienten per Telefoninterview und bei 118 Patienten per Papierformat. Bei der zweiten Befragung wurde der jeweils andere Modus genutzt mit 94 Patienten in der Papierbefragung und 118 Patienten im Telefoninterview. Hiermit ergab sich eine Gesamtkohorte zur Auswertung des Vergleichs zwischen Telefoninterview zum Papierformat von 212 Patienten (98 % der Studiengruppe). Als dritter Modus erfolgte bei 204 der 216 Patienten (94 %) das erneute Telefoninterview zur Erhebung der Daten für die zeitliche Reliabilität. Als vierter Modus wurde von 63 Patienten (29 %) der elektronische Fragebogen bearbeitet, um anschließend Telefoninterview und Papierformat mit dem elektronischen Fragebogen zu vergleichen (s. mod. Abb. 1). Die durchschnittlichen Befragungsintervalle betragen 18 ± 13 Tage bei der zweimaligen telefonischen Befragung, 12 ± 12 Tage zwischen Telefoninterview- und Papierformat, 20 ± 13 Tage zwischen Papierformat- und elektronischem Fragebogen und 17 ± 11 Tage zwischen Telefoninterview- und elektronischem Fragebogen. Hinsichtlich soziodemografischer Merkmale und okulärer oder systemischer Begleiterkrankungen fanden sich keine signifikanten Unterschiede zwischen der Gruppe, die sich zunächst zur selbstadministrierten Papierform entschied und der Gruppe, welche zunächst telefonisch befragt wurde (s. Supp. Tab. 1 (Terheyden et al., 2024)).

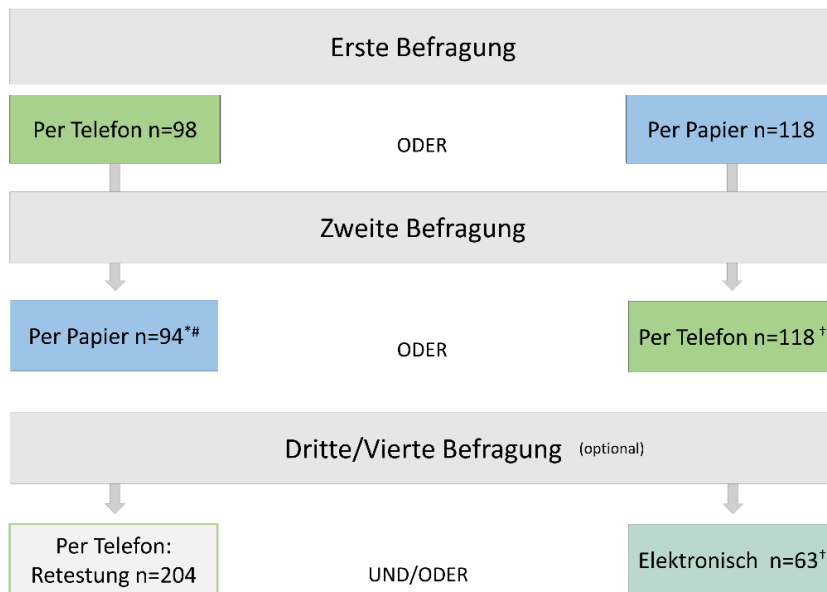


Abb. 1, modifiziert nach Abb. 2 aus Veröffentlichung (Terheyden et al., 2024) : Die Abbildung zeigt den Ablauf der Befragungen, wobei pro Studienteilnehmer bis zur vier Befragungen erfolgt sind. * Vier Teilnehmer haben die Befragung im Papiermodus aufgrund ausgeprägter Visuseinschränkung nicht durchgeführt. # Ein Teilnehmer hat den Papiermodus als dritte Befragungsversion durchgeführt. † Zwei Teilnehmer haben diesen Modus als dritte Befragungsversion durchgeführt.

1.3.3 Psychometrische Auswertung des IVI

Bei der psychometrischen Auswertung des Impact of Vision Impairment – Fragebogens mittels Rasch-Analyse lagen der Personenreliabilitätskoeffizient (PR) und der Personentrennungsindex (PSI) zur Beurteilung der internen Konsistenz für alle Berechnungen innerhalb der Kategorien und bis auf eine Ausnahme bei der telefonischen Administration der emotionalen Subskala des Fragebogens immer über 2,0 (PSI) beziehungsweise 0,8 (PR) (s. Tab. 2 (Terheyden et al., 2024)). Es ergaben sich zudem keine Veränderungen bezüglich der Ordnung innerhalb der Antwortkategorien. Bei der Überprüfung der Infit/Oufit-Statistik fielen bei drei Items moderat abweichende Werte zwischen 1,40 und 1,56 auf, welche jedoch die Gesamtberechnungen nicht signifikant beeinflussten (Wright und Linacre, 1994) (s. Tab. 2 (Terheyden et al., 2024)). Um auch geringfügige Abweichungen auszuschließen, berechneten wir die Ergebnisse nach Entfernung der Items 1, 14 und 21 erneut, wobei sich jedoch keine Veränderungen zeigten, sodass die Items für die weitere Analyse beibehalten wurden. Bei den telefonischen Befragungen fanden sich leichte Abweichungen vom ersten Kontrast der

Residuen bei der funktionalen Subskala der Papierbefragung und beiden Subskalen des elektronischen Fragebogens, jedoch zeigte sich kein Nachweis von Multidimensionalität. Die PCA für den ersten Faktor erklärte >55 % bis 73 % der Varianz in Abhängigkeit von dem jeweiligen Befragungsmodus. In der weiteren Analyse fanden sich Cluster, die an die Subgruppenverteilung des englischsprachigen IVI (Reading, Mobility, Emotional) erinnerten (Lamoureux et al., 2007c). Bei genauerer Betrachtung widersprach die diskontinuierliche Korrelation von $>+0,7$ jedoch der Existenz einer zusätzlichen Dimension. Es zeigte sich eine deutliche Überschreitung von 1 bei der Differenz zwischen Personen- und Itemmittelwerten (s. Tab. 2 (Terheyden et al., 2024)). Bei der Überprüfung einer mit dem Alter oder dem Geschlecht korrelierenden Antworttendenz ergaben sich bei drei Items Abweichungen: Geschlecht beim IVI_F, telefonisches Interview, ICC 0,997 [95-% Konfidenzintervall: 0,996;0,998]; Geschlecht beim IVI_F, Papierfragebogen, ICC 0,996 [0,995;0,997]; Alter beim IVI_E, elektronischer Fragebogen, ICC 0,979 [0,847; 0,993]. In diesen Fällen entfernten wir die Items und die Personenmesswerte wurden erneut berechnet. Hierdurch wurden jedoch keine Veränderungen hervorgerufen und die Items wurden beibehalten (s. Tab. 2 (Terheyden et al., 2024)). Die im direkten Vergleich höchsten Personenmesswerte fanden sich bei der telefonischen Interviewbefragung (s. Supp. Tab. 2 (Terheyden et al., 2024)).

1.3.4 Reliabilitätsanalyse

Zusammenfassend waren die Intraklassenkorrelationskoeffizienten (ICC) mit Werten zwischen 0,853 und 0,939 für alle verglichenen Gruppen als „ausgezeichnet“ einzuordnen (Cicchetti und Sparrow, 1981) (s. Tab. 3 (Terheyden et al., 2024)). Es fanden sich ähnliche mittlere Differenzen bei allen Vergleichen und eine systematische Verzerrung war in der Bland-Altman-Analyse nicht festzustellen (s. mod. Abb. 2). Eine detaillierte Ergebnisdarstellung findet sich in den folgenden Absätzen geordnet nach dem jeweiligen Vergleichsmodus (Retest / Vergleich der Befragungsmodi).

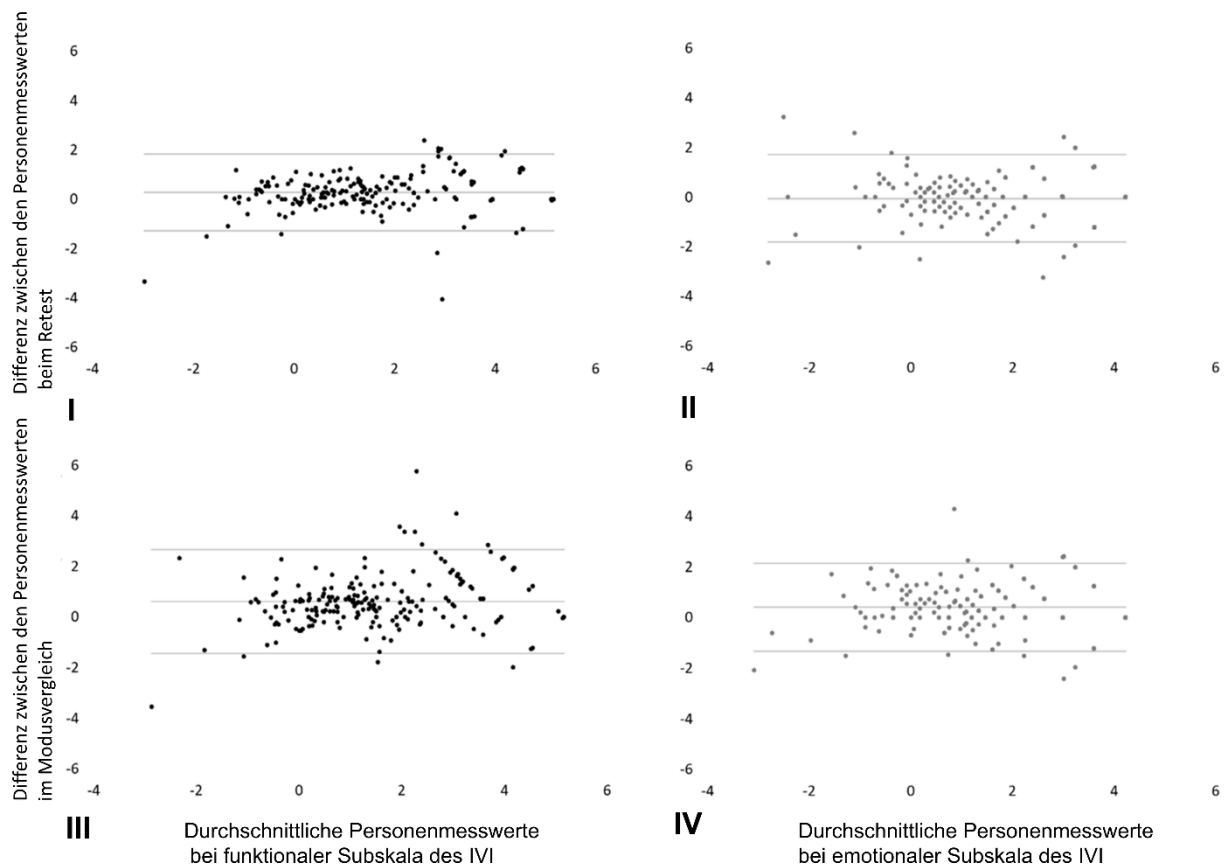


Abb. 2, modifiziert nach Abb. 3 aus Veröffentlichung (Terheyden et al., 2024). Dargestellt sind vier Bland-Altman-Diagramme (links funktionale Subskalen, rechts emotionale Subskalen) zur Veranschaulichung des Verteilungsmusters der mittleren Personenmesswerte und der Differenz der Personenmesswerte im Vergleich zwischen 1. und 2. Befragung = Retest (I, II) sowie im Vergleich zwischen Telefon- und Papierbefragung (III, IV).

1.3.4.1 Retest

Die zweimalige telefonische Befragung entsprach mit ICC-Werten von $>0,9$ der ab $ICC=0,75$ beginnenden Kategorie „ausgezeichnet“ (Cicchetti und Sparrow, 1981). Dennoch fand sich bei der Auswertung der funktionalen Subskala eine leicht positive mittlere Differenz (s. mod. Abb. 2 und Tab. 3 (Terheyden et al., 2024)). Dies entspricht geringfügig höheren Personenmesswerten beim ersten Telefonat und somit der Angabe einer etwas besseren sehbezogenen Lebensqualität (VRQoL) bei der telefonischen Erstbefragung. Bei der emotionalen Subskala zeigte sich hingegen eine negative mittlere Differenz, welche damit der Angabe einer etwas besseren emotionalen VRQoL beim zweiten

Telefonat entspricht (s. mod. Abb. 2 und Tab. 3 (Terheyden et al., 2024)). Der Wiederholungskoeffizient (CoR) zeigte mit 21,2 % der Teilskala einen registrierbaren Unterschied zwischen Erst- und Zweitbefragung in der emotionale Subskala an (s. Tab. 3 (Terheyden et al., 2024)). Bei der Aufarbeitung der Werte mittels Deming-Regression fanden sich ähnliche Regressionslinien ohne signifikante Unterschiede. Lediglich bei der Betrachtung höherer Personenmesswerte zeigten sich bei der funktionalen Subskala etwas höhere, jedoch weiterhin nicht signifikante Ergebnisse für das erste Telefonat (s. Tab. 3 (Terheyden et al., 2024)).

1.3.4.2 Vergleich der Befragungsmodi

Der ICC lag für alle Auswertungen bei $>0,8$ und entspricht damit der höchsten Kategorie „ausgezeichnet“ (Cicchetti und Sparrow, 1981). Unter Berücksichtigung der 95%-Konfidenzintervalle ist jedoch der „Telefon“-„Papier“-Vergleich in der funktionellen Subskala getrennt zu betrachten, da dann eine Einteilung in die nächsttiefere Kategorie „gut“ erfolgt (Cicchetti und Sparrow, 1981; Koo und Li, 2016). Insgesamt lag die mittlere Differenz bei allen Vergleichen in einem niedrigen Bereich von 0,033 bis 0,611 (s. Tab. 3 (Terheyden et al., 2024)). Der höchste Wiederholbarkeitskoeffizient (CoR) zeigte sich mit einem Wert von über 2 bei dem Vergleich der telefonischen Befragung gegenüber dem Papierfragebogen, wobei dies 21 % der Spannweite der Subskala entspricht (s. Tab. 3 (Terheyden et al., 2024)). Die Bland-Altman-Analyse ergab eine positive mittlere Differenz bei der Telefonbefragung gegenüber Papierfragebogen und dem elektronischen Fragebogen und beschrieb damit eine geringere Angabe von Beeinträchtigungen im „Telefonmodus“ (s. mod. Abb. 2 und Tab. 3 (Terheyden et al., 2024)). Bei dem Vergleich des Papierfragebogens mit dem elektronischen Fragebogen ergab sich in beiden Subskalen eine negative mittlere Differenz und somit eine geringere Angabe von Beeinträchtigungen bei der elektronischen Version (s. Tab. 3 (Terheyden et al., 2024)).

Bei der Auswertung der Personenmesswerte mittels Deming-Regressionsanalyse fanden sich beim Vergleich der Administrationsmodi einige signifikante Abweichungen (s. Tab. 3 (Terheyden et al., 2024)). Die telefonische Befragung führte im Vergleich zur papiergebundenen und elektronischen Befragung zu signifikant höheren Werten der funktionalen Subskala entsprechend einer höheren VRQoL (s. Tab. 3 (Terheyden et al.,

2024). Die papiergebundene Befragung zeigte im Vergleich zur telefonischen und elektronischen Befragung signifikant niedrigere Werten bei der emotionalen IVI-Subskala und dementsprechend eine niedrigere VRQoL (s. Tab. 3 und Supp. Tab. 2 (Terheyden et al., 2024)).

1.3.4.3 Störfaktoren-Analyse

Im Kontext der Datenerhebung wurden potentielle Einflussfaktoren auf die Ergebnisgenauigkeit identifiziert. Hierbei fanden sich unter Anwendung der Bonferroni-Methode für multiples Testen weder für die bestkorrigierte Sehschärfe, noch für die selbstberichteten Hörprobleme oder psychische Beeinträchtigungen signifikante Auffälligkeiten im Mittelwertvergleich. Auch hinsichtlich des Befragungsintervalls und des primären Befragungsmodus fanden sich mit korrigiertem $p \geq 1$ keine signifikanten Zusammenhänge bezüglich der zuvor beschriebenen Unterschiede bei der IVI-Subskalenauswertung (s. Supp. Tab. 3 (Terheyden et al., 2024)).

1.4 Diskussion

In dieser Studie zeigte sich, dass beim IVI sowohl die wiederholten Befragungen als auch die intermodalen Vergleiche zwischen Telefonmodus, Papiermodus und elektronischem Modus kaum zu Abweichungen in der Beantwortung führten. Die ausführliche Analyse erfolgt aufgeteilt nach den zwei Haupthypothesen.

Hypothese 1: Es bestehen keine Unterschiede zwischen den Antworten der Erst- und der Folgebefragung des IVI.

Insgesamt zeigten sich in unserer Studie beim Retest-Vergleich des IVI mit Intraklassenkorrelationskoeffizienten (ICCs) von $>0,9$ hohe Übereinstimmungen, welche damit zu vorherigen IVI- und PROM-Studien passen. In bisherigen IVI-Retest-Studien wurden kleinere Kohorten mit 20 bis 60 Probanden ausgewertet (Marakis et al., 2020; Ratanasukon et al., 2016; Weih et al., 2002). Im Vergleich zum 32-Item IVI von Weih et al. mit einem Wert von 0,88 und mittleren Differenzen von 6,1 % des Subskalenbereichs gegenüber in unserer Studie $\leq 2,8$ % des Skalenbereichs fand sich in unserer Studie eine etwas höhere Übereinstimmung der Ergebnisse bezüglich der unterschiedlichen Befragungszeitpunkte (Weih et al., 2002). Andere Auswertungen des IVI wie die der

thailändischen oder die griechischen Validierung ergaben mit hohen ICCs von $\geq 0,90$ sehr ähnliche Ergebnisse wie in unserer Studie (Marakis et al., 2020; Ratanasukon et al., 2016). Zwischen den verschiedenen PROM-Instrumenten der Augenheilkunde bestehen große Ähnlichkeiten. Daher bieten beispielsweise der Vergleich zum NEI-VFQ 25, einem Fragebogen zur Abbildung alltagsrelevanter Sehbeeinträchtigungen, sowie zum VF-14, einem Fragebogen zu sehbezogenen Alltagsaktivitäten, eine weitere Möglichkeit die IVI-Studienergebnisse im wissenschaftlichen Kontext einzuordnen. Es zeigt sich, dass die dort präsentierten ICCs von $> 0,75$ nach der Einteilung von Cicchetti et al. ebenfalls unter die Kategorie „exzellent“ fallen und sich damit die IVI-Ergebnisse in den Auswertungsbereich anderer bekannter PROMs in der Augenheilkunde einreihen (Cassard et al., 1995; Cicchetti und Sparrow, 1981; Labiris et al., 2008; Nichols et al., 2002). Insgesamt lässt sich daraus schlussfolgern, dass sich die Beantwortung der Fragen zur visuellen Lebensqualität beim IVI vom Zeitpunkt und der Frequenz der Befragungen kaum beeinflussen lässt.

Hypothese 2: Es bestehen keine Unterschiede zwischen den Antworten bei der Befragung mittels unterschiedlicher Administrationsmodi (Telefonische Version, Papierversion, Elektronische Version).

Die Auswertung der verschiedenen Administrationsmodi zeigte insgesamt keinen signifikant höheren Einfluss auf die IVI-Subskalenwerte als die einer Abweichung durch die Durchführung einer erneuten Befragung (s. Tab. 3 (Terheyden et al., 2024)). Die auffälligsten Abweichungen zeigten sich mit Intraklassenkorrelationskoeffizienten von $< 0,9$ bei dem Vergleich von Telefonmodus und Papierformat mit leicht höheren Bewertungen der Lebensqualität bei der Interview-Befragung. Insbesondere bei diesem Vergleich sollte daher ein Blick auf die Deming-Regression und die Mittelwertdifferenzen zur Gesamteinschätzung der Ergebnisse erfolgen. Der Ordinatenabschnitt und die Steigung bei der Deming-Regressionsanalyse lagen nur knapp oberhalb des Signifikanzlevels und im Hinblick auf die Mittelwertdifferenzen fanden sich diese ebenfalls in einem nur gering erhöhten Bereich von $\leq 6,2$ % bei dem funktionellen IVI-Abschnitt bis zu unter $\leq 4,9$ % im Bereich der emotionalen Subskala, sodass abgesehen von der statistischen Auffälligkeit eine Gleichwertigkeit der Befragungsmodi hinsichtlich des klinischen Alltages zu erwarten ist (s. Tab. 3 (Terheyden et al., 2024)). Wenn man diese

geringfügige Abweichung in der Literatur sucht, findet sich in der Vision Core Measure (VCM1) - Studie, einem anderen PROM aus der Augenheilkunde, ebenfalls eine Abweichung hinsichtlich einer besseren VRQoL bei der Beantwortung im Interviewmodus im Vergleich zu der selbstadministrierten Befragungen mittels Papierfragebogen (Frost et al., 2001; Wolffsohn et al., 2000). Die Ursache könnte in der Untererfassung von Alltagsschwierigkeiten und emotionalen Beeinträchtigungen bei der Befragung im Interviewmodus liegen, da eine Beantwortung im Sinne einer sozialen Erwünschtheit bei direktem Kontakt eine größere Rolle spielt (Rickwood und Coleman-Rose, 2023; Aquilino, 1998).

Zur Einordnung unserer Ergebnisse zwischen Papier- und elektronischem Fragebogen erfolgte der Vergleich mit einer Übersichtsarbeit zu PROMs im Allgemeinen. Diese Übersicht stellt einen Prozentwert von 5 % des Skalenbereichs als Grenze hinsichtlich relevanter Abweichungen auf, wobei unsere Ergebnisse lediglich einen Unterschied von $\leq 2,3$ % in Bezug auf die jeweilige Subskala ergaben und somit von einer nicht relevanten Abweichung auszugehen ist (Gwaltney et al., 2008; Terheyden et al., 2024). Hinsichtlich ophthalmologischer PROMs fanden sich wenige systematische Vergleiche zwischen webbasierter und papierbasierter Befragung, die berichteten aber ähnliche Reliabilitätswerte wie in unserer Studie (Clayton et al., 2013; Terheyden et al., 2022). Eine modernere Variante im Vergleich zur Webbasierten Befragung stellt das App-Format dar. Beim Ocular Surface Disease Index (OSDI) zur Beurteilung einer Keratoconjunctivitis sicca wurde der Vergleich zwischen App-Format und Papierbefragung durchgeführt und ergab ebenfalls eine sehr gute Übereinstimmung (Nagino et al., 2023), sodass auch eine Weiterentwicklung des IVI diesbezüglich möglich scheint.

Zum Vergleich unserer Studienergebnisse mit vorherigen Untersuchungen des IVI liegt unseres Wissens nur eine Studie mit 31 Probanden vor, bei welcher der Interviewmodus mit der Selbstadministration im Papierformat verglichen wurde (Weih et al., 2002). Die Studie lässt sich zudem nur bedingt vergleichen, da es sich um die 32-Item Version handelt und die Auswertung nicht mittels psychometrischer Rasch-Analyse, sondern anhand von Summenscores erfolgte. Im Vergleich der mittleren Differenzen von $\leq 6,2$ % in unserer Studie gegenüber $\leq 9,5$ % in der IVI-Analyse von 2002 zeigt unsere Studie eine höhere Übereinstimmung zwischen den Befragungsmodi auf, als von Weih et al. berichtet

(Terheyden et al., 2024; Weih et al., 2002). Der bei uns auffällige leichte Unterschied zwischen Interviewer- und Selbstadministration zeigte sich ebenfalls in der zuvor genannten Studie beim Vergleich der Intraklassenkorrelationskoeffizienten mit 0,91 bei der wiederholten Interviewer-geführten Befragung gegenüber 0,84 bei der mit Interview und anschließend mit Selbstadministration durchgeführten Befragung (Weih et al., 2002). Die Bewertung dieser leichten Abweichung erfolgte vergleichbar mit unseren Daten dahingehend, dass davon ausgegangen wurde, dass bei Veränderungen der Ergebnisse nicht die Administrationsmethode verantwortlich ist, sondern eine veränderte Beurteilung der visusbezogenen Lebensqualität des Patienten vorliegt (Weih et al., 2002).

Aufgrund unseres stark ökonomisierten Gesundheitswesens und den begrenzten zeitlichen Ressourcen sehen wir unsere Ergebnisse als unterstützendes Element zur zunehmenden Nutzung selbstadministrierbarer Fragebögen und dabei insbesondere bei der Einführung elektronischer Befragungsinstrumente in der Ophthalmologie, da durch automatisierte Prozesse eine Reduktion des Verwaltungsaufwandes sowie der Personalkosten erreicht werden können (Evans und Mathur, 2005; McHorney et al., 1994; Lavalée et al., 2016). Die Ophthalmologie ist ein Fachbereich, der von zeitlichen Ökonomisierungen besonders profitiert, da ein sehr hohes Patientenaufkommen mit Bedarf vieler Kontrolluntersuchungen besteht und daher bei kurzen Behandlungseinheiten das bereits erwähnte paternalistische Prinzip weiterhin verhältnismäßig ausgeprägt umgesetzt wird (Framme et al., 2022; Bouaziz et al., 2022). Insbesondere sehr ausholende Anamnesegespräche könnten durch Nutzung des IVI strukturierter und damit zeitlich effektiver durchgeführt werden, da mittels PROMS persönliche Erfahrungen und die Wahrnehmungen zur Lebensqualität bereits vorher eingeordnet werden können (Lavalée et al., 2016). Daher ist es nicht nur von methodischer Relevanz, dass die Fragebögen Veränderungen in der Lebensqualität registrieren, denn so wird sowohl dem Arzt als auch dem Patient nachvollziehbar dargestellt in welchen Bereichen ein Therapieansatz erfolgreich ist und wo möglicherweise Verbesserungsbedarf besteht (Lavalée et al., 2016).

Unsere Studie zeichnet sich insbesondere durch die Anwendbarkeit der Ergebnisse auf eine relativ große Patientengruppe aus. Dies liegt einerseits an der hohen Anzahl an Studienteilnehmern, sowie andererseits an den offen gewählten Einschlusskriterien. Dies

bietet die Möglichkeit einer gewissen Verallgemeinerung unserer Ergebnisse auf die Gesamtheit ophthalmologischer Patienten. Hinsichtlich der gewählten Auswertungsverfahren wurden mit der Rasch-Analyse latente Eigenschaftsmodelle genutzt, welche die direkte Anwendung parametrischer Testverfahren erlauben und damit aufgrund der statistisch höheren Aussagekraft der herkömmlichen Auswertung auf Ordinalskalenniveau vorzuziehen sind (Braithwaite et al., 2019).

Die zuvor dargestellten Stärken der Studie stellen andererseits zum Teil auch Schwachpunkte der Studie dar, denn durch die gewählten Einschlusskriterien wurden bei der Rekrutierung Patienten mit jeglicher Form von chronischer Augenerkrankung eingeschlossen und es wurde keine Visusobergrenze festgelegt, sodass auch nicht sehbeeinträchtigte Patienten an der Studie teilnahmen. Hierdurch wurde in der psychometrischen Analyse - wie auch bereits in früheren Studien des IVI - eine Differenz zwischen Personen- und Itemmittelwerten festgestellt mit der Konsequenz, dass unsere Probandenkohorte keine ideale Zielgruppe für den IVI darstellt (Lamoureux et al., 2007a; Goldstein et al., 2018). Ein weiterer Kritikpunkt stellt die fehlende Randomisierung bezüglich des ersten Befragungsmodus dar, wobei jedoch in einer anschließenden Analyse keine Antwortdifferenzen diesbezüglich festgestellt werden konnten.

1.5 Zusammenfassung

In unserer Studie wurde der Impact of Vision Impairment – Fragebogen als bekannter PROM zur sehbezogenen Lebensqualität an einer Kohorte von 216 Patienten hinsichtlich zweier Fragestellungen untersucht. Einerseits erfolgte die Analyse der Frage nach der Zuverlässigkeit der wiederholten Befragung (Retestrelabilität). Andererseits wurde ein Vergleich von drei unterschiedlichen Administrationsmodi herausgearbeitet und die Frage nach der Austauschbarkeit der verschiedenen Befragungsmodi beantwortet. Die Auswertungen fanden alle unter Verwendung moderner statistischer Ansätze wie Rasch-basierter Personenmesswerte sowie Intraklassenkorrelation und Deming-Regression statt. Es bestätigte sich im Einklang mit vorherigen Untersuchungen bei nur geringen und nicht signifikanten Unterschieden in unserer Studie eine ausgezeichnete Retestrelabilität des IVI. Bei dem intermodalen Vergleich von Telefon-, Papier- und elektronischem Modus ließen sich insgesamt nur geringfügigen Abweichungen feststellen. Insbesondere die

Befragung mittels Papierfragebogen zeigte im Vergleich zum Telefoninterview eine leicht verringerte VRQoL, jedoch unterhalb eines klinisch relevanten Bereiches.

Zusammenfassend zeigen die Studienergebnisse, dass der IVI sowohl in der wiederholten Befragung als auch in der Nutzung verschiedener Administrationsmodi einen zuverlässigen PROM darstellt. Zudem lässt sich festhalten, dass unsere Ergebnisse weder durch die Art der ersten Befragung, noch durch die Sehschärfe des besser sehenden Auges, durch eine Höreinschränkung der Teilnehmer, durch psychische Beeinträchtigungen oder durch die Zeit zwischen den Befragungen signifikant beeinflusst wurden.

Abschließend ist davon auszugehen, dass die Studienergebnisse den Einsatz des IVI im klinischen Alltag und in der ophthalmologischen Forschung durch die Möglichkeit einer individuell angepassten Befragungsmethode und insbesondere die Verwendung elektronischer Fragebögen erleichtern. Die Verwendung des IVI als Instrument zur Erfassung der Lebensqualität im Therapieverlauf ist hinsichtlich der Beurteilung eines Therapieeffektes sehr hilfreich und unterstützt zudem die Möglichkeit als Patient seine persönlichen Wahrnehmungen strukturiert mitzuteilen und damit auch im Sinne der partizipativen Entscheidungsfindung einfacher einen Einfluss auf die weitere Therapieplanung zu nehmen. Um den IVI in diesem Kontext umfänglicher zu nutzen, wäre beispielweise eine Integration des Fragebogens in den Anamneseprozess mit Erhebung des Erstbefundes denkbar, da dadurch von vornherein stärker patientenzentrierte Diagnostik- und Therapieentscheidungen ermöglicht würden. Eine damit verbundene Reduktion des organisatorischen und ökonomischen Aufwandes müsste durch weitere Studien ermittelt werden, könnte letztendlich jedoch die Patientenposition relevant stärken.

1.6 Literaturverzeichnis der deutschen Zusammenfassung

- Aquilino WS. Effects of Interview Mode on Measuring Depression in Younger Adults. *J Off Stat* 1998; 14: 15–29
- Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013; 346: f167
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310
- Boer MR de, Moll AC, Vet HCW de, ..., van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt* 2004; 24: 257–273
- Bond TG, Fox CM. Applying the Rasch model: Fundamental Measurement in the Human Sciences. 1st ed. New York: Lawrence Erlbaum Associates Publishers 2001
- Boone WJ, Noltemeyer A. Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education* 2017; 4: 202
- Bouaziz M, Cheng T, Minuti A, ..., Barmettler A. Shared Decision Making in Ophthalmology: A Scoping Review. *Am J Ophthalmol* 2022; 237: 146–153
- Braithwaite T, Calvert M, Gray A, ..., Denniston AK. The use of patient-reported outcome research in modern ophthalmology: impact on clinical trials and routine clinical practice. *Patient Relat Outcome Meas* 2019; 10: 9–24
- Cassard SD, Patrick DL, Damiano AM, ..., Steinberg EP. Reproducibility and responsiveness of the VF-14. An index of functional impairment in patients with cataracts. *Arch Ophthalmol* 1995; 113: 1508–1513
- Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic* 1981; 86: 127–137
- Ciccione L, Dehaene S. Can humans perform mental regression on a graph? Accuracy and bias in the perception of scatterplots. *Cogn Psychol* 2021; 128: 101406
- Clayton JA, Eydelman M, Vitale S, ..., Ferris F. Web-based versus paper administration of common ophthalmic questionnaires: comparison of subscale scores. *Ophthalmology* 2013; 120: 2151–2159
- Dean S, Mathers JM, Calvert M, ..., Denniston AK. "The patient is speaking": discovering the patient voice in ophthalmology. *Br J Ophthalmol* 2017; 101: 700–708
- Deming WE. Statistical adjustment of data: J. Wiley & Sons, Incorporated 1943
- Denniston AK, Kyte D, Calvert M, ..., Burr JM. An introduction to patient-reported outcome measures in ophthalmic research. *Eye (Lond)* 2014; 28: 637–645

Evans JR, Mathur A. The value of online surveys. *Internet Research* 2005; 15: 195–219

Finger RP, Fenwick E, Marella M, ..., Lamoureux EL. The impact of vision impairment on vision-specific quality of life in Germany. *Invest Ophthalmol Vis Sci* 2011; 52: 3613–3619

Finger RP, Guymer RH, Gillies MC, ..., Keeffe JE. The impact of anti-vascular endothelial growth factor treatment on quality of life in neovascular age-related macular degeneration. *Ophthalmology* 2014; 121: 1246–1251

Framme C, Dittberner M, Rohwer-Mensching K, ..., Volkmann I. Leistungs- und Kostenkalkulation für eine universitäre, augenheilkundliche Hochschulambulanz. *Ophthalmologe* 2022; 119: 46–54

Frost MH, Reeve BB, Liepa AM, ..., Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 2007; 10: 94–105

Frost NA, Sparrow JM, Hopper CD, ..., Peters TJ. Reliability of the VCM1 Questionnaire when administered by post and by telephone. *Ophthalmic Epidemiol* 2001; 8: 1–11

Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)* 2015; 25: 141–151

Goldstein JE, Fenwick E, Finger RP, ..., Massof R. Calibrating the Impact of Vision Impairment (IVI): Creation of a Sample-Independent Visual Function Measure for Patient-Centered Outcomes Research. *Transl Vis Sci Technol* 2018; 7: 38

Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health* 2008; 11: 322–333

Hanmer J, Hays RD, Fryback DG. Mode of administration is important in US national estimates of health-related quality of life. *Med Care* 2007; 45: 1171–1179

Hassell JB, Lamoureux EL, Keeffe JE. Impact of age related macular degeneration on quality of life. *Br J Ophthalmol* 2006; 90: 593–596

Hassell JB, Weih LM, Keeffe JE. A measure of handicap for low vision rehabilitation: the impact of vision impairment profile. *Clin Exp Ophthalmol* 2000; 28: 156–161

Herzog AR, Rodgers WL. Interviewing Older Adults: Mode Comparison Using Data from a Face-to-Face Survey and a Telephone Resurvey. *Public Opin Q* 1988; 52: 84–99

Hochstim JR. A Critical Comparison of Three Strategies of Collecting Data from Households. *J Am Stat Assoc* 1967; 62: 976

Joosten EAG, DeFuentes-Merillas L, Weert GH de, ..., Jong CAJ de. Systematic review of the effects of shared decision-making on patient satisfaction, treatment adherence and health status. *Psychother Psychosom* 2008; 77: 219–226

Kanellopoulos AJ. The impact of keratoconus treatment with the Athens Protocol (partial topography-guided photorefractive keratectomy combined with higher-fluence corneal collagen cross-linking) on quality of life: a long-term study. *Clin Ophthalmol* 2019; 13: 795–803

Keeffe JE, McCarty CA, Hassell JB, ..., Gilbert AG. Description and measurement of handicap caused by vision impairment. *Aust N Z J Ophthalmol* 1999; 27: 184–186

Knauer C, Pfeiffer N. The value of vision. *Graefes Arch Clin Exp Ophthalmol* 2008; 246: 477–482

Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016; 15: 155–163

Labiris G, Katsanos A, Fanariotis M, ..., Tsironi E. Psychometric properties of the Greek version of the NEI-VFQ 25. *BMC Ophthalmol* 2008; 8: 4

Lamoureux EL, Ferraro JG, Pallant JF, ..., Keeffe JE. Are standard instruments valid for the assessment of quality of life and symptoms in glaucoma? *Optom Vis Sci* 2007a; 84: 789–796

Lamoureux EL, Pallant JF, Pesudovs K, ..., Keeffe JE. The Impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci* 2006; 47: 4732–4741

Lamoureux EL, Pallant JF, Pesudovs K, ..., Keeffe JE. The effectiveness of low-vision rehabilitation on participation in daily living and quality of life. *Invest Ophthalmol Vis Sci* 2007b; 48: 1476–1482

Lamoureux EL, Pallant JF, Pesudovs K, ..., Keeffe JE. The impact of vision impairment questionnaire: an assessment of its domain structure using confirmatory factor analysis and rasch analysis. *Invest Ophthalmol Vis Sci* 2007c; 48: 1001–1006

Lavallee DC, Chenok KE, Love RM, ..., Franklin PD. Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. *Health Aff (Millwood)* 2016; 35: 575–582

Linacre JM. *A User's Guide to Winsteps/Ministep*. Chicago, IL: Mesa Press 2006

Loukanova S, Bridges J. Empowerment in medicine: An analysis of publication trends 1980–2005. *Open Medicine* 2008; 3: 105–110

Luu W, Kalloniatis M, Bartley E, ..., Ly A. A holistic model of low vision care for improving vision-related quality of life. *Clin Exp Optom* 2020; 103: 733–741

Marakis TP, Koutsandrea C, Poulou MS. The impact of vision impairment on vision-related quality of life of patients with neovascular age-related macular degeneration. *Eur J Ophthalmol* 2020; 32: 481–490

Massof RW, Rubin GS. Visual Function Assessment Questionnaires. *Surv Ophthalmol* 2001; 45: 531–548

- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1996; 1: 30–46
- McHorney CA, Kosinski M, Ware JE. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. *Med Care* 1994; 32: 551–567
- Nagino K, Okumura Y, Akasaki Y, ..., Inomata T. Smartphone App-Based and Paper-Based Patient-Reported Outcomes Using a Disease-Specific Questionnaire for Dry Eye Disease: Randomized Crossover Equivalence Study. *Journal of medical Internet research* 2023; 25: e42638
- Nichols KK, Mitchell GL, Zadnik K. Performance and repeatability of the NEI-VFQ-25 in patients with dry eye. *Cornea* 2002; 21: 578–583
- Pondorfer SG, Terheyden JH, Heinemann M, ..., Finger RP. Association of Vision-related Quality of Life with Visual Function in Age-Related Macular Degeneration. *Sci Rep* 2019; 9: 15326
- Ratanasukon M, Tongsomboon J, Bhurayanontachai P, ..., Jirarattanasopa P. The Impact of Vision Impairment (IVI) Questionnaire; Validation of the Thai-Version and the Implementation on Vision-Related Quality of Life in Thai Rural Community. *PLoS ONE* 2016; 11: e0155509
- Rickwood DJ, Coleman-Rose CL. The effect of survey administration mode on youth mental health measures: Social desirability bias and sensitive questions. *Heliyon* 2023; 9: e20131
- Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. *Med Decis Making* 2015; 35: 114–131
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420–428
- Spatz ES, Krumholz HM, Moulton BW. Prime Time for Shared Decision Making. *JAMA* 2017; 317: 1309–1310
- Tan JCK, Nguyen V, Fenwick E, ..., Watson SL. Vision-Related Quality of Life in Keratoconus: A Save Sight Keratoconus Registry Study. *Cornea* 2019; 38: 600–604
- Tapp H, Shade L, Mahabaleshwarkar R, ..., Dulin MF. Results from a pragmatic prospective cohort study: Shared decision making improves outcomes for children with asthma. *J Asthma* 2017; 54: 392–402
- Terheyden JH, Mekschat L, Ost RAD, ..., Finger RP. Interviewer Administration Corresponds to Self-Administration of the Vision Impairment in Low Luminance (VILL) Questionnaire. *Transl Vis Sci Technol* 2022; 11: 21

Terheyden JH, Ost RAD, Behning C, ..., Finger RP. Evaluation of the test-retest and inter-mode comparability of the Impact of Vision Impairment questionnaire in people with chronic eye diseases. *Graefes Arch Clin Exp Ophthalmol* 2024

Varma R, Wu J, Chong K, ..., Hays RD. Impact of severity and bilaterality of visual impairment on health-related quality of life. *Ophthalmology* 2006; 113: 1846–1853

Vaz S, Falkmer T, Passmore AE, ..., Andreou P. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS ONE* 2013; 8: e73990

Vet HCW de, Bouter LM, Bezemer PD, ..., Beurskens AJHM. Reproducibility and responsiveness of evaluative outcome measures. *Int J Technol Assess Health Care* 2001; 17: 479–487

Weih LM, Hassell JB, Keeffe J. Assessment of the Impact of Vision Impairment. *Invest Ophthalmol Vis Sci* 2002; 43: 927–935

Wolffsohn JS, Cochrane AL, Watt NA. Implementation methods for vision related quality of life questionnaires. *Br J Ophthalmol* 2000; 84: 1035–1040

World Health Organization. 1998. Programme on mental health. Verfügbar unter https://iris.who.int/bitstream/handle/10665/77932/WHO_HIS_HSI_Rev.2012.03_eng.pdf?sequence=1 (Zugriffsdatum: 01.07.2024)

Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 1989; 70: 857–860

Wright BD, Linacre JM. 1994. Reasonable mean-square fit values. *Rasch Measurement Transactions*. Verfügbar unter <https://www.rasch.org/rmt/rmt83b.htm> (Zugriffsdatum: 27.08.2024)

Wright BD, Masters GN. *Rating Scale Analysis*. Chicago, Ill.: Mesa Pr 1982

Xu RH, Cheung AWL, Wong ELY. The relationship between shared decision-making and health-related quality of life among patients in Hong Kong SAR, China. *Int J Qual Health Care* 2017; 29: 534–540

2. Veröffentlichungen

Dieser Publikationsdissertation liegt die folgende, unabhängig begutachtete Veröffentlichung zugrunde:

Terheyden JH, Ost RAD, Behning C, Mekschrat L, Bildik G, Wintergerst MWM, Holz FG, Finger RP. Evaluation of the test-retest and inter-mode comparability of the Impact of Vision Impairment questionnaire in people with chronic eye diseases. Graefe's Archive 2024; 262(6):1933-1943

<https://doi.org/10.1007/s00417-023-06334-4>

3. Erklärung zum Eigenanteil

Die Erarbeitung und Planung des Studienablaufes erfolgte in enger Zusammenarbeit mit Herrn Prof. Dr. Dr. Finger. Die Datenerhebung führte ich maßgeblich alleine durch, einzelne organisatorische Teilschritte zu Beginn der Studie wurden durch Frau Mekschat und Frau Bildik koordiniert. Die Auswertung wurde durch mich durchgeführt, wobei die Deming-Regressionsanalysen durch Frau Behning erfolgten und Herr PD Dr. Terheyden mich bei Auswahl der Analyseverfahren unterstützte. Die Interpretation und Einordnung der Ergebnisse in den wissenschaftlichen Kontext erfolgte durch mich unter Supervision von Herrn PD Dr. Terheyden und meinem Doktorvater Herrn Prof. Dr. Dr. Finger.

Ich versichere, die Dissertationsschrift selbständig verfasst zu haben und keine weiteren als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

4. Danksagung

Ich möchte mich herzlich bei allen Menschen bedanken, die mich bei der Erstellung dieser Arbeit unterstützt haben.

An erster Stelle gilt mein Dank meinem Doktorvater Herrn Prof. Dr. Dr. Robert Finger, der mir die Möglichkeit zu dieser Dissertation eröffnet und mich bei der Erarbeitung maßgeblich unterstützt hat. Mein besonderer Dank gilt auch Herrn PD Dr. Jan Terheyden für seine sehr engagierte und immer motivierende Betreuung. Meinen Eltern, meinem Bruder und meinem Freund danke ich sehr für ihr offenes Ohr und die vielen inspirierenden Gespräche, die mich während der Promotionszeit, aber auch während des gesamten Studiums, begleitet haben.

5. Publikationen (PDF-Version)

Graefe's Archive for Clinical and Experimental Ophthalmology (2024) 262:1933–1943
<https://doi.org/10.1007/s00417-023-06334-4>

LOW VISION



Evaluation of the test–retest and inter-mode comparability of the Impact of Vision Impairment questionnaire in people with chronic eye diseases

Jan Henrik Terheyden¹ · Reglind A. D. Ost¹ · Charlotte Behning² · Liza Mekschrat¹ · Gamze Bildik¹ · Maximilian W. M. Wintergerst¹ · Frank G. Holz¹ · Robert P. Finger³

Received: 14 June 2023 / Revised: 3 November 2023 / Accepted: 30 November 2023 / Published online: 5 January 2024
© The Author(s) 2024

Abstract

Purpose The main objective of this study is to assess the test–retest and inter-administration mode reliability of the Impact of Vision Impairment profile (IVI), a common patient-reported outcome measure (PROM) for people with chronic eye diseases.

Methods The IVI was administered to adult patients with stable, chronic eye diseases two to four times per participant (average intervals between administrations 12 to 20 days; maximum two phone interviews, paper administration, electronic administration) by two trained interviewers. Rasch models were fit to the data. Intra-class correlation coefficients (ICCs), mean differences and Cronbach's alpha between test–retest administrations (two phone interviews) and inter-mode comparisons were calculated.

Results Two hundred-sixteen patients (mean age 67 ± 12 years, 40% male) were included in the study. The IVI met all psychometric requirements of the Rasch model, and the division into the domains of functional items (IVI_F) and emotional items (IVI_E) corresponded to the German validation study. ICCs (all for IVI_F and IVI_E, respectively) for the retest administrations were 0.938 and 0.912, and 0.853 and 0.893 for inter-mode comparisons phone/paper, 0.939 and 0.930 for phone/electronic, and 0.937 and 0.920 for paper/electronic (all $p < 0.01$). Mean differences (all for IVI_F and IVI_E, respectively) for the retest administrations were 2.8% and 0.7% and ranged from 2.0% to 6.2% and from 0.4 % to 4.9% between administration modes. Cronbach's alpha ranged from 0.886 to 0.944 for retest and inter-mode comparisons.

Conclusion Due to the high test–retest reliability and the almost equally high comparability of different modes of administration of the IVI, the study endorses its use as a robust PROM to capture vision-related quality of life. Our results further support the use of the IVI as an endpoint in clinical trials and may simplify implementing it in both clinical trials or real-world evidence generation by offering multiple administration modes with high reliability.

Keywords Patient-reported outcome · Reliability · Impact of Vision Impairment scale · Quality of life

Jan Henrik Terheyden and Reglind A.D. Ost contributed equally to the work.

Extended author information available on the last page of the article

Key messages

What is known:

- The Impact of Vision Impairment profile is a commonly used instrument for measuring vision-related quality of life
- Prior to this study, test-retest reliability data were only available from small cohorts

What is new:

- Our results suggest a high test-retest reliability in the current version of the IVI in a sample of >200 participants
- The data of our study demonstrate high comparability between three different modes of questionnaire administration

Introduction

Patient-reported outcome measures (PROMs) continue to gain importance in both routine eye care provision and research [1–4]. PROMs consist of several questions assessing the patient's health from their perspective [4]. One of the most common concepts measured by PROMs in ophthalmology is vision-related quality of life (VRQoL). It is known to be compromised by even mild forms of visual impairment and is therefore considered as a useful tool to gain additional information to visual function testing [5–7]. Typical PROMs in ophthalmology to assess VRQoL or visual functioning are the Impact of Vision Impairment (IVI) profile, the National Eye Institute Visual Functioning Questionnaire-25 (NEI-VFQ-25), and the Visual Function Index-14 (VF14) [4, 8, 9]. Similar to psychophysical function tests, PROM instruments require careful validation [10, 11]. One distinguishes between qualitative validation steps, where, e.g., content validity is assessed during the development phase of a questionnaire and quantitative validation steps [12]. The latter include an assessment of reliability and, e.g., construct validity of a questionnaire by statistical techniques such as exploratory factor analysis, to measure if the questions measure the construct they are intended to measure [12]. Similarly, reliability is assessed using quantitative metrics such as Cronbach's alpha or Rasch model person reliability to assess the internal consistency of the scale, and methods comparing two assessments with the same patient at different time points to examine the test–retest reliability of the scale [12].

Both the NEI-VFQ-25 and the VF14 have undergone test–retest reliability testing in large cohorts (NEI-VFQ-25 $n = 186$; VF-14 $n = 383$, which showed a high agreement between repeated administrations of both instruments [13, 14]. The IVI is widely used in rehabilitation and treatment studies [9, 11, 15–20], but to date, data on test–retest reliability have been generated in three studies that have included comparatively small subgroups of participants ($n \leq 60$ participants per study, total 102 test–retest participants in all available studies) [21–23]. Regarding this limited evidence,

we identified the need to reevaluate test–retest reliability in a larger cohort.

Another important aspect of questionnaire studies besides repeatability over different time points is the repeatability across different administration modes. Many questionnaire studies are still conducted with self-administered paper questionnaires or with interviews. However, due to numerous advantages of electronic modes of administration, such as higher and faster response rates, lower costs, and simplified data analysis, these have become more popular in research [24]. To our knowledge, a systematic comparison of electronic, paper-based, and interviewer-based modes has not been conducted in any of the studies mentioned above. One single study has compared self-administration of the IVI using a paper form and interviewer administration in a cohort with 31 participants, which is a similarly small sample size compared to the test–retest reliability studies described [23].

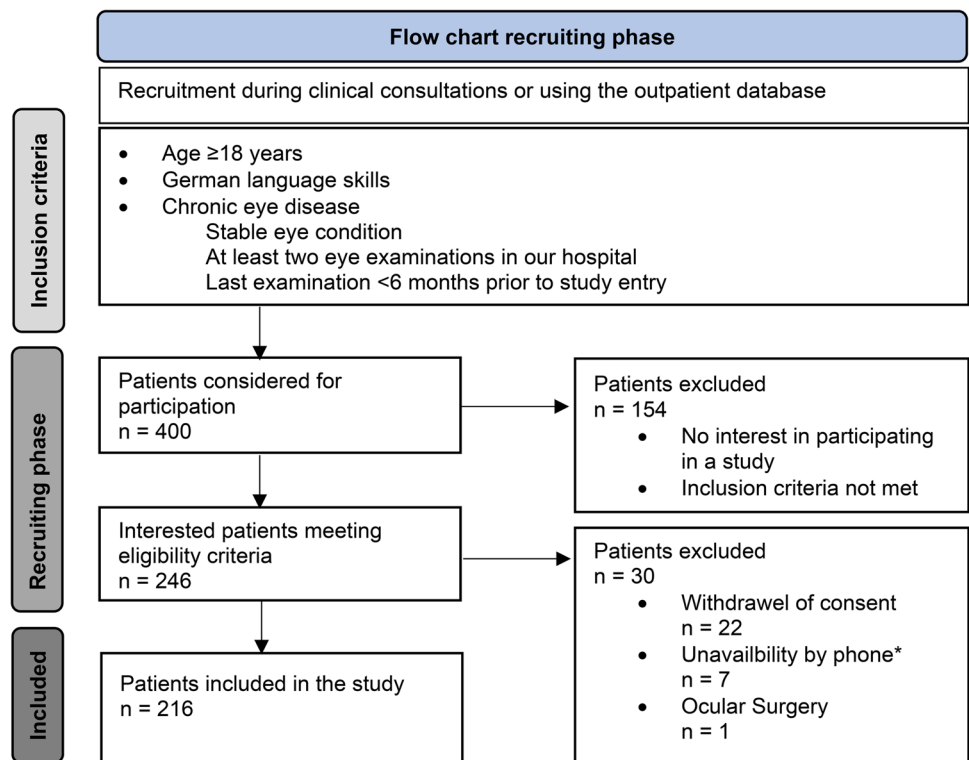
To fill this gap, our objective was to investigate the test–retest reliability and inter-mode comparability of the IVI concerning paper, phone, and electronic administration modes, hypothesizing that retest assessments are comparable but interviewer administration yielding higher IVI scores [25, 26].

Methods

Recruitment

Our prospective study was carried out at the Department of Ophthalmology, University Hospital Bonn, Germany, from April 2020 to December 2020. Ethical approval was obtained from the Institutional Review Board of the University Hospital Bonn (reference number: 130/16). Patients were recruited during clinical consultations and from an outpatient database. Inclusion criteria were chronic eye conditions, age ≥ 18 years and sufficient German language skills. We only included individuals with stable chronic

Fig. 1 Flow chart of patient recruitment. *At least five phone call attempts



eye condition for at least two previous examinations at our hospital, the latter being no more than 6 months prior to study entry. Exclusion criteria were any acute-onset eye diseases <6 weeks, any eye surgery or interventions during or <6 weeks prior to study participation except intravitreal injections. The size of the sample was chosen according to international recommendations [27]. Written informed consent was obtained from all participants of the study. The study protocol followed the principles of the Declaration of Helsinki (Fig. 1).

The Impact of Vision Impairment questionnaire

The IVI questionnaire consists of 28 items to detect different aspects of VRQoL in patients with visual impairment [28]. It was originally developed to evaluate whether patients are limited in their daily lives due to reduced vision and may need rehabilitation [29, 30]. As indicated in its validation study, the German IVI includes four response options for items 1–13 and five response options for items 14–28, starting with “not at all” and ending with “very often” to evaluate the items’ influence on VRQoL. The fifth option “don’t do this for other reasons” was treated as missing in our analysis. The questionnaire is divided into two subscales: “Functional IVI” (items 1–20) and “Emotional IVI” (items 21–28) [5]. In accordance with the original IVI, the German version was initially developed for interview administration [5, 29].

Questionnaire administration

The study included two to four administrations of the IVI for each participant (Fig. 2). The initial mode of administration depended on the participant’s preference. The required response interval between administrations was ≤ 10 weeks. During the repeated administrations, the previous questionnaire responses were unavailable to both the interviewer and the participant. The interviewer administrations were conducted by two trained interviewers who instructed the participants according to recommended PROM administration guidelines [31].

In addition to the administration of the IVI, participants self-administered structured questionnaires on socio-demographical data and medical history. Diagnoses and visual acuity data from the last visit were obtained from the patient files. Regarding missing data, the questionnaire was excluded from the study if responses were available for fewer than 50% of items.

Psychometric validation of the IVI in our population

We performed Rasch analysis, which transforms the ordinal Likert scales into pseudo-interval-level scales (expressed in logits) and thus allows for parametric statistical analysis using commercial software (Winsteps software, ver. 3.92.1.2; Winsteps, Chicago, IL), employing the Andrich

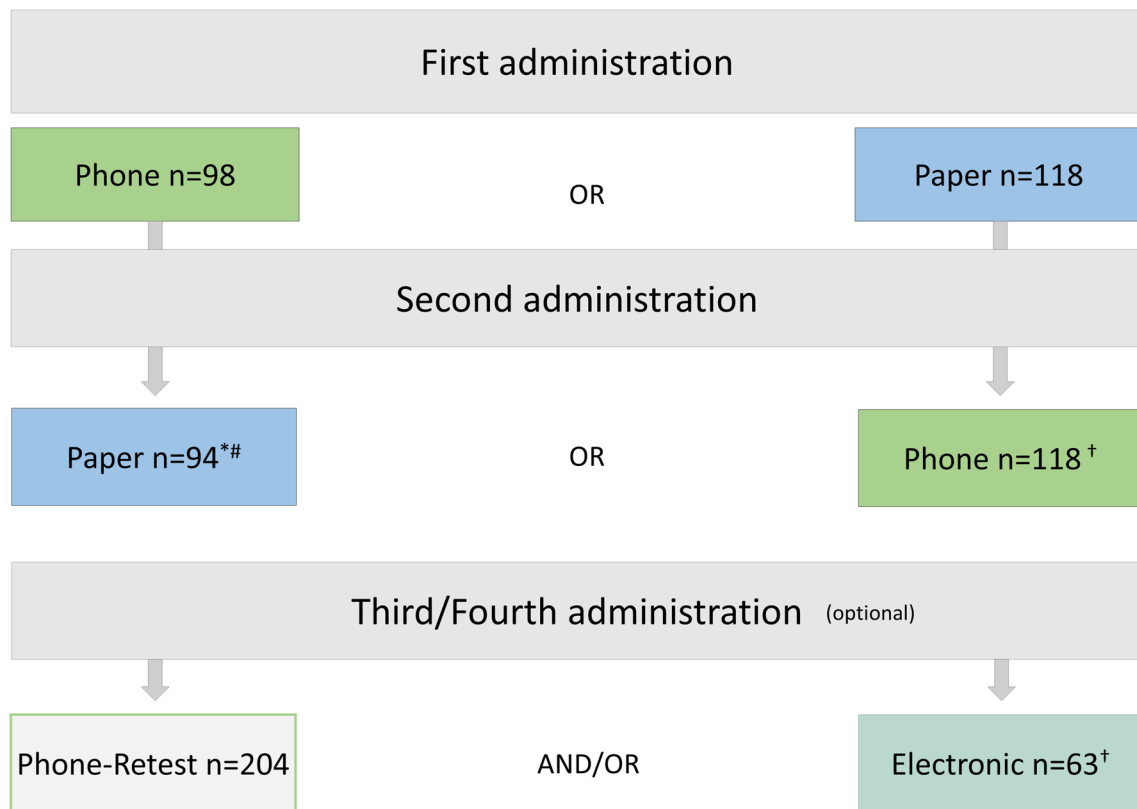


Fig. 2 Order of questionnaire administrations. Up to four sets of IVI questionnaire data were collected per participants. Interviewer administrations were performed remotely (phone) and self-administrations were performed using paper or electronic forms. *Omitted in 4 par-

ticipants due to vision impairment; #Performed as third mode by one participant; †Performed as third mode by two participants (after paper and electronic)

Rating Scale model [32, 33]. As an instance of latent trait models, the application of a Rasch model minimizes the impact of single missing responses since the measured construct is largely independent from individual missing values [34]. We anchored the person measures across modes of administration to the Rasch model based on the first interviewer administration to enable direct comparability of the results within our study, as well as in relation to other studies of the IVI [5, 35, 36]. Apart from this, we also analyzed item measures of Rasch models without this anchoring step, by Pearson correlation coefficients. After processing the questionnaire data, we inverted the person measures of the IVI to facilitate interpretation by assigning lower person measures to participants with more severe vision impairment.

Based on the Rasch models, we assessed threshold ordering, person reliability, person separation, item fit, and differential item functioning (DIF). Response category threshold ordering was assessed to verify whether the category structure belonging to the items is matching [28]. A person separation index (PSI) above 2.0 was considered as good and values between 1.5 and 2.0 as

acceptable, and a person reliability (PR) above 0.8 was deemed high and between 0.6 and 0.8 deemed acceptable [37–39]. Unidimensionality concerning the underlying construct was assessed using the infit and outfit mean square standardized residuals (MNSQ) statistic, with values up to 1.4 being reasonable for a rating scale survey [40]. The targeting of the instrument was identified by calculating the difference between person and item means, with values > 1.0 logits representing mistargeting. Lastly, differential item functioning (DIF) was assessed for each item to measure differences between subgroups concerning sex and age and was noticeable for values over 0.64 [41]. To determine the significance of deviating DIF values, a comparison of the person measures with and without retention of the conspicuous values was performed using intra-class correlation coefficients (ICCs).

Statistical analysis

The main outcomes of our study were ICCs and mean differences of repeated administration. We investigated

(1) *t*-retest and (2) inter-mode reliability using two-way mixed effect ICCs with absolute agreement as ICCs are the preferred assessment for settings where measures are repeated [42–44]. ICCs were interpreted according to the categories suggested by Cicchetti and Sparrow (poor $ICC < 0.40$; fair, $0.40 \leq ICC \leq 0.59$; good, $0.60 \leq ICC \leq 0.74$; excellent, $0.75 \leq ICC \leq 1.00$) [45]. Furthermore, we displayed the mean of inter-measurement differences ± 1.96 standard deviations as limits of agreement in Bland–Altman plots [46]. The coefficient of repeatability (CoR, i.e., an indicator of absolute reliability) was calculated as 1.96 standard deviations of the mean test–retest difference [47, 48]. Cronbach's alpha, a conventional test theory metric of reliability, was also calculated for the test–retest reliability assessment, based on person measures [12]. Deming regression analysis was performed to compare person measures between different administrations [49]. In contrast to classical linear regression, Deming regression considers the errors in both variables included in the respective regression model [50]. Intercept and slope were evaluated for significant deviations from 0 and 1, respectively [49]. To rule out any systematic bias, we compared the distribution of personal characteristics of participants who selected interviewer assessment and participants who selected self-assessment as the initial mode of administration, using the *t* test and the chi-squared test. In addition, we investigated which influence the initial mode of administration, visual acuity, self-reported hearing difficulties, self-reported psychiatric diseases, and the interval between administrations had on the mean differences between test–retest administration and phone-paper administration, using the Mann–Whitney *U* test.

Statistical analyses were performed with SPSS (Version 27 software, IBM SPSS Inc., Armonk, NY) and R (version 4.1.2). *P* values < 0.05 were considered statistically significant with correcting for multiple testing as necessary.

Results

Two hundred sixteen out of 246 total participants completed at least two administrations of the IVI and were included in the analysis. The reasons for drop-out were withdrawal of consent ($n = 22$), unavailability of participants by phone ($n = 7$), and ocular surgery ($n = 1$). The majority of participants (160 individuals, 74%) was above 60 years of age (Table 1). The most frequent ocular conditions in our cohort were vitreoretinal diseases, glaucoma and cataract (Supplementary Table 1). A total of 204 individuals were included in the test–retest assessment. The availability of administration modes for the inter-mode assessment was 216 participants (100%) for interviewer administrations, 212 (98%) for self-administrations using pen-and-paper,

Table 1 Characteristics of the sample

	<i>n</i> (%)
	All ($n = 216$)
Age	
Mean age [years] \pm SD	67 \pm 12
Sex	
Female (%)	129 (59.7)
Male (%)	87 (40.3)
Visual acuity (logMAR)	
Hearing difficulties	0.27 \pm 0.14
Yes (%)	59 (27.3)
No (%)	156 (72.2)
Missing (%)	1 (0.4)
Education	
Elementary school (%)	76 (35.2)
Secondary school (%)	78 (36.1)
High school (%)	22 (10.2)
University with graduation (%)	36 (16.7)
Missing (%)	4 (1.9)
Employment status	
Working (%)	56 (25.9)
Unemployed (%)	22 (10.2)
Retired (%)	130 (60.2)
Missing (%)	8 (3.7)
Living situation	
Alone (%)	64 (29.6)
With others (%)	148 (68.5)
Missing (%)	4 (1.9)
Marital status	
Married (%)	127 (58.8)
Widowed (%)	36 (16.7)
Divorced (%)	32 (14.8)
Unmarried (%)	21 (9.7)

and 63 (29%) for electronic self-administrations. The mean administration intervals were 18 ± 13 days for the test–retest assessments, 12 ± 12 days between phone and paper administrations, 20 ± 13 days between paper and electronic administrations, and 17 ± 11 days between phone and electronic administrations. The number of individuals who chose self-administration as the initial mode of administration ($n = 118$) was higher than the number of participants who chose to be interviewed first ($n = 98$) but no differences between these groups in terms of socio-demographic characteristics, ocular or systemic concurrent disease were present (all $p \geq 0.087$, Supplementary Table 1). Questionnaires with at least half of the responses missing were excluded from the study ($n = 5$). The number of single missing values amounted to 1.5%, or 0.01% of all answers, excluding the values of the answer option “Don't do this for other reasons”, which was considered missing in our analysis.

Table 2 Fit parameters of the phone, paper and electronic administration for the functional and emotional IVI

Parameters	Rasch model	Phone administration		Paper administration		Electronic administration	
		IVI_F	IVI_E	IVI_F	IVI_E	IVI_F	IVI_E
Threshold ordering, <i>n</i>	0	0	0	0	0	0	0
Misfitting items, <i>n</i>	0	0	item21^a	0	0	item1^a; item14^a	0
PSI	> 2.0 (1.5)	2.55	1.76	3.32	2.71	3.33	2.19
PR	> 0.8 (0.6)	0.87	0.76	0.92	0.88	0.92	0.83
Difference in person and item mean	< 1	1.91	2.60	1.56	2.87	2.38	3.97
DIF	< 0.64						
Age (≤ 60 ; > 60)		None	None	None	None	None	item28^b
Sex (female; male)		item1^b	None	item1^b; item9^b; item18^b	None	None	None

DIF, Differential item functioning; *IVI_F*, functional subscale of the IVI; *IVI_E*, emotional subscale of the IVI; *PR*, person reliability; *PSI*, person separation index

Bold values represent misfit to the Rasch model

^aValues are not degrading the measurement system [40]

^bNo influence on measurements ($ICC \geq 0.98$)

Rasch model fit

All response category thresholds were ordered. Three items demonstrated misfit (infit MNSQ 1.45, item 1; 1.56, item 14; 1.40, item 21), but the values were below a level that degrades the measurement system, and removal did not improve fit statistics, so we retained the items [40]. PR and PSI indicated adequate internal consistency (Table 2). The difference in person and item mean showed poor person-item targeting in our cohort. Three items were indicative of DIF by sex and one item by age group, but person measures were unchanged after removal, and the items could therefore be retained (*IVI_F*, phone administration, ICC 0.997, 95%-CI [0.996; 0.998]; *IVI_F*, paper administration, ICC 0.996, 95%-CI [0.995; 0.997]; *IVI_E*, electronic administration, ICC 0.979, 95%-CI [0.847; 0.993]). The resulting person measures were highest for the interview administrated IVI (Supplementary Table 2). Pearson correlation coefficients between item measures were calculated to validate the chosen anchoring method. They were 0.975 [0.946; 0.988] for the test and retest-phone administrations, 0.878 [0.751; 0.942] for phone and paper administration, and 0.877 [0.748; 0.942] for phone and electronic administration.

Test-retest and inter-mode reliability

ICCs were excellent, both for test-retest and inter-mode comparisons (Table 3). Mean differences between test-retest and inter-mode comparisons were comparable, and the Bland-Altman analysis did not indicate any systematic bias (Fig. 3). Deming regression revealed that absolute differences of test-retest assessments were not significantly

different from each other when the same mode of administration was used. Phone administration compared to paper and electronic administration resulted in significantly higher functional IVI subscale scores (i.e., higher VRQoL), and paper administration compared to phone and electronic administration yielded significantly lower emotional IVI subscale scores (i.e., lower VRQoL) when comparing the Deming regression intercepts (Table 3; Supplementary Table 2).

Sensitivity analysis

To validate the above findings, we investigated relationships between mean differences of the test-retest or inter-mode comparisons and five potential confounders, correcting for multiple testing with the Bonferroni method. Neither best-corrected visual acuity (corrected $p \geq 0.546$), nor psychiatric diseases (corrected $p \geq 1.0$), self-reported hearing difficulties (corrected $p \geq 1.0$), administration interval (corrected $p \geq 1.0$), nor the initial mode of administration (corrected $p \geq 1.0$) were significantly associated with the IVI subscale score differences (Supplementary Table 3).

Discussion

The purpose of this study was to evaluate test-retest reliability and inter-mode reliability. Test-retest differences were small and non-significant, and inter-mode variations were in a similar, small range. We found a small but significant reduction in VRQoL scores when questionnaires were self-administered using paper forms compared to phone administration. Due

Table 3 Reliability metrics of the test–retest and inter-mode assessments for the functional and emotional IVI subscales

		Test–retest		Inter-mode	
		<i>Phone–phone</i> (<i>n</i> = 204)	<i>Phone–paper</i> (<i>n</i> = 212)	<i>Phone–electronic</i> (<i>n</i> = 63)	<i>Paper–electronic</i> (<i>n</i> = 63)
IVI_F	ICC (average values)	0.938	0.853	0.939	0.937
	95%-CI	[0.909; 0.957]	[0.693; 0.917]	[0.896; 0.964]	[0.894; 0.962]
	Cronbach's α	0.94	0.89	0.94	0.94
	Mean difference (% subscale range)	0.27 (2.8)	0.61 (6.2)	0.24 (2.5)	–0.20 (2.0)
	CoR (% subscale range)	1.55 (15.8)	2.06 (21.0)	1.61 (16.5)	1.53 (15.6)
	Deming, intercept	–0.10	–0.29	–0.27	–0.18
	95%-CI	[–0.27; 0.07]	[–0.53; –0.06]*	[–0.54; –0.01]*	[–0.43; 0.07]
	Deming, slope	0.91	0.83	1.02	1.24
	95%-CI	[0.82; 0.99]*	[0.72; 0.94]*	[0.86; 1.17]	[1.06; 1.42]*
	IVI_E	ICC (average values)	0.912	0.893	0.930
95%-CI		[0.884; 0.933]	[0.816; 0.933]	[0.885; 0.958]	[0.866; 0.952]
Cronbach's α		0.91	0.91	0.93	0.92
Mean difference (% subscale range)		–0.06 (0.7)	0.41 (4.9)	0.03 (0.4)	–0.20 (2.3)
CoR (% subscale range)		1.77 (21.2)	1.75 (20.9)	1.42 (16.9)	1.40 (16.8)
Deming, intercept		0.01	–0.38	0.25	0.35
95%-CI		[–0.27; 0.28]	[–0.60; –0.16]*	[–0.17; 0.67]	[0.16; 0.54]*
Deming, slope		1.03	0.99	0.83	0.89
95%-CI		[0.90; 1.16]	[0.88; 1.09]	[0.62; 1.04]	[0.76; 1.03]

ICC, Intra-class correlation coefficient; *IVI_F*, functional subscale; *IVI_E*, emotional subscale; *ICC*, intra-class correlation coefficient; *CoR*, coefficient of repeatability

*Regression intercept significantly different from 0 or slope significantly different from 1

to the excellent ICC values (*IVI_F* 0.853; *IVI_E* 0.893) and low mean differences (*IVI_F* 6% of subscale range, *IVI_E* 5% of subscale range), we classified these deviations in the statistically but not clinically relevant range [45]. Overall, our data demonstrate that the IVI questionnaire is highly reliable, independent of repeated assessment or the mode of administration, also supported by the strong associations of the item measures across different modes. Our results were not affected by the initial mode of administration, participants' visual or hearing impairment, psychiatric diseases, or the time between IVI administrations.

The test–retest reliability of the IVI was largely comparable to what has been reported previously in smaller cohorts (*n* = 20 to 60) [21–23]. ICCs between IVI test–retest scores in our sample (0.94 for *IVI_F* and 0.91 for *IVI_E*) were minimally higher compared to the values reported by the original 32-item IVI by Weih et al. (0.88 for IVI total sum score) and mean differences were lower in our study ($\leq 2.8\%$ of subscale range versus 6.1% of scale range) [23]. Test–retest subgroup analyses in the Greek and Thai validations of the IVI reported similar results to our study (ICCs ≥ 0.90 , Cronbach's $\alpha > 0.75$) [21, 22]. Other commonly used ophthalmic PROMs, such as the NEI-VFQ-25 or the VF-14, achieve

similarly ICCs (0.57–0.91), which places the IVI in the series of reliable PROMs in ophthalmology [13, 14, 51].

The implementation of different modes of administration did not have a relevantly higher impact on the IVI subscale scores than performing a retest administration in our study. As indicated above, the inter-mode reliability has only been investigated in one study (*n* = 31) that compared self-administration to interview-administration, using IVI-32 sum scores which are no longer recommended to be used [23]. In contrast to Weih et al. we have investigated the inter-mode reliability of the IVI using more state-of-the-art statistical approaches including Rasch-based person measures. Our results suggested a higher agreement between modes of administration than previously reported (mean differences $\leq 6.2\%$ in this study versus $\leq 9.5\%$ in the previous analysis) [23]. Additionally, our findings of the electronic questionnaire with excellent ICCs ≥ 0.920 further support the implementation of electronic PROMs in ophthalmology (Table 3) [24].

We found a trend of higher reported vision-related quality of life in interviews compared to pen and paper assessments, which is known from other PROMs such as the vision core measure (VCM1) study by Frost et al. [25]. In our dataset,

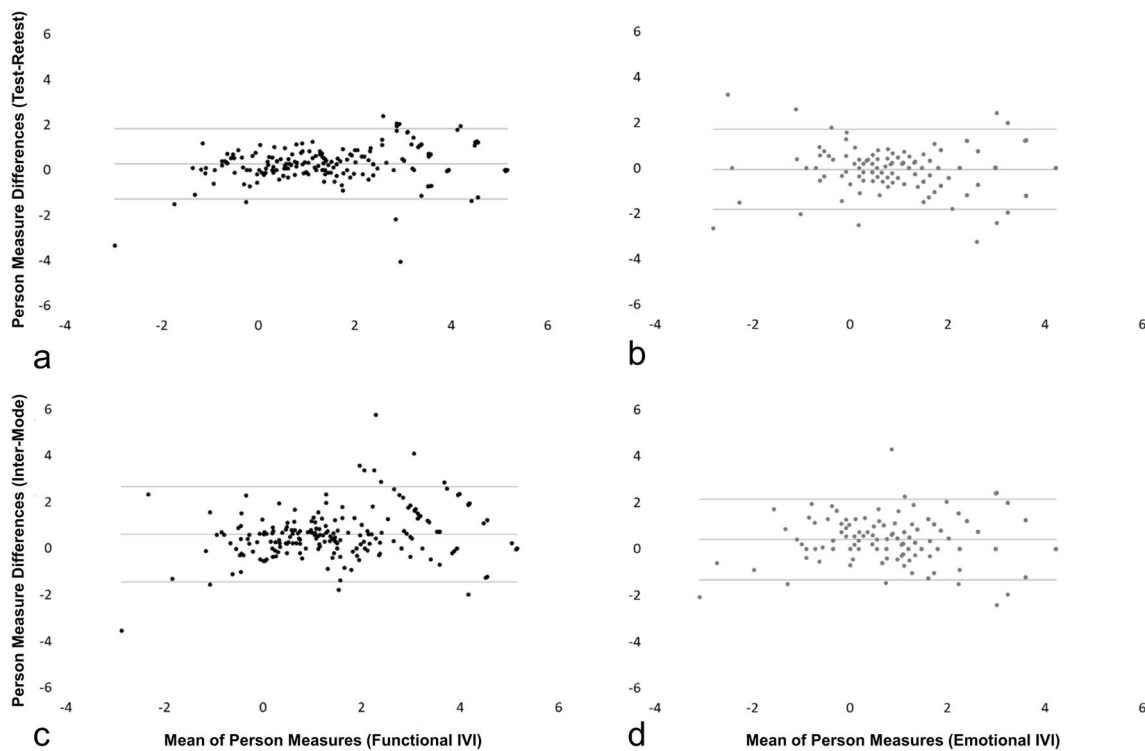


Fig. 3 Bland–Altman plots of test–retest assessments (phone administration) of the functional (a) and emotional (b) subscales and of inter-mode comparability assessments (phone and paper administra-

tion) of the functional (c) and emotional (d) IVI subscales display the distributions of retest differences in person measures (higher vision-related quality of life corresponds to higher person measures)

mean differences between different modes of administration were small ($\leq 6.2\%$ of functional subscale range; $\leq 4.9\%$ of emotional subscale range), further supporting the equivalence between the investigated administration modes for the IVI. When comparing paper to electronic self-administration, mean differences were $\leq 2.3\%$ of the respective subscale range. This is in line with previous research, which has identified 5% of the scale range as the cutoff value for most studies comparing paper to electronic administrations of various PROMs [52]. Only few other studies have systematically compared the use of different modes of administration in ophthalmic PROMs, and reported inter-mode reliabilities were in an overall similar range to the results of our study [53, 54].

The main limitation of our study is the targeting of the sample which is a result of including individuals with several chronic eye diseases but not only visually impaired individuals [55]. The problem of inadequate person-item targeting is known from previous studies of the IVI [56]. We did not randomize participants by initial mode of administration since most participants asked to start with pen and paper administration. Yet, a post hoc analysis did not reveal any differences based on the initial mode of administration. An additional aspect not investigated in our study is how the participants' health literacy may

modify the reliability of the IVI, which should be further investigated in future studies [57].

Our study's main strengths include its large sample size in which both test–retest and inter-mode reliability of the IVI profile have been tested, the heterogeneity of the sample making the results more likely to be generalizable to a population accessing eye care services, and the use of latent trait models which have several known advantages over sum scoring in psychometric assessment [58, 59].

In conclusion, the IVI questionnaire demonstrated excellent test–retest reliability and our data suggest the use of interview, paper-based, or electronic modes of administration to be comparable in our study cohort. The results may facilitate implementing the IVI in clinical routine and research because of a reduced administration burden.

Abbreviations *CoR*: Coefficient of repeatability; *DIF*: Differential item functioning; *ICC*: Intra-class correlation coefficient; *IVI*: Impact of Vision Impairment profile; *IVI_F*: Functional subscale of the Impact of Vision Impairment profile; *IVI_E*: Emotional subscale of the Impact of Vision Impairment profile; *MNSQ*: Mean square standardized residuals; *PR*: Person reliability; *PROMs*: Patient-reported outcome measures; *PSI*: Person separation index; *VRQoL*: Vision-related quality of life

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00417-023-06334-4>.

Author contribution JHT, RADO, and RPF designed the study. RADO, LM, and GB collected the data. JHT, RADO, CB, RPF, MWMW, and FGH analyzed and interpreted the data. JHT and RADO were main contributors in writing the manuscript. All authors read, substantively revised, and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical approval Ethical approval was obtained from the Institutional Review Board of the University Hospital Bonn (reference number: 130/16). The study protocol followed the principles of the Declaration of Helsinki.

Informed consent Written informed consent was obtained from all participants of the study.

Consent for publication Not applicable.

Competing interests JHT: Heidelberg Engineering, Optos, Zeiss, CenterVue, Novartis, Okko; RADO: Heidelberg Engineering, Optos, Zeiss, CenterVue; CB: none; LM: Heidelberg Engineering, Optos, Zeiss, CenterVue; GB: Heidelberg Engineering, Optos, Zeiss, CenterVue; MWMW: Heidelberg Engineering (F, R), Optos (F), Carl Zeiss Meditec (F), CenterVue (F), Heine Optotechnik GmbH (C, R, F), Berlin-Chemie AG (F), Novartis Pharma GmbH (R, F), D-Eye Srl (F), Eyenuk Inc. (F), ASKIN & CO GmbH (R), DigiSight Technologies (R); FGH: Acucela, Allergan, Apellis, Bayer, Boehringer-Ingelheim, Bioeq/Formycon, CenterVue, Ellex, Roche/Genentech, Geuder, Grayburg Vision, Heidelberg Engineering, Kanghong, LinBioscience, NightStarX, Novartis, Optos, Pixium Vision, Oxurion, Stealth Bio-Therapeutics, Zeiss; RPF: Bayer, Ellex, Novartis, Novartis, Opthea, Alimera, Santhera, Roche/Genentech, CentreVue, Zeiss.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.




References

- Loukanova S, Bridges J (2008) Empowerment in medicine: an analysis of publication trends 1980–2005. *Open Med* 3:105–110. <https://doi.org/10.2478/s11536-007-0066-7>
- Nelson EC, Eftimovska E, Lind C et al (2015) Patient reported outcome measures in practice. *BMJ* 350:g7818. <https://doi.org/10.1136/bmj.g7818>
- Dean S, Mathers JM, Calvert M et al (2017) “The patient is speaking”: discovering the patient voice in ophthalmology. *Br J Ophthalmol* 101:700–708. <https://doi.org/10.1136/bjophthalmol-2016-309955>
- Braithwaite T, Calvert M, Gray A et al (2019) The use of patient-reported outcome research in modern ophthalmology: impact on clinical trials and routine clinical practice. *Patient Relat Outcome Meas* 10:9–24. <https://doi.org/10.2147/PROM.S162802>
- Finger RP, Fenwick E, Marella M et al (2011) The Impact of Vision Impairment on vision-specific quality of life in Germany. *Invest Ophthalmol Vis Sci* 52:3613–3619. <https://doi.org/10.1167/iovs.10-7127>
- Varma R, Wu J, Chong K et al (2006) Impact of severity and bilaterality of visual impairment on health-related quality of life. *Ophthalmology* 113:1846–1853. <https://doi.org/10.1016/j.ophtha.2006.04.028>
- Knauer C, Pfeiffer N (2008) The value of vision. *Graefes Arch Clin Exp Ophthalmol* 246:477–482. <https://doi.org/10.1007/s00417-007-0668-4>
- de Boer MR, Pluijm SMF, Lips P et al (2004) Different aspects of visual impairment as risk factors for falls and fractures in older men and women. *J Bone Miner Res* 19:1539–1547. <https://doi.org/10.1359/JBMR.040504>
- Luu W, Kalloniatis M, Bartley E et al (2020) A holistic model of low vision care for improving vision-related quality of life. *Clin Exp Optom* 103:733–741. <https://doi.org/10.1111/cxo.13054>
- Massof RW, Rubin GS (2001) Visual function assessment questionnaires. *Surv Ophthalmol* 45:531–548. [https://doi.org/10.1016/S0039-6257\(01\)00194-1](https://doi.org/10.1016/S0039-6257(01)00194-1)
- Lamoureux EL, Pallant JF, Pesudovs K et al (2007) The effectiveness of low-vision rehabilitation on participation in daily living and quality of life. *Invest Ophthalmol Vis Sci* 48:1476–1482. <https://doi.org/10.1167/IOVS.06-0610>
- Frost MH, Reeve BB, Liepa AM et al (2007) What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 10(Suppl 2):94–105. <https://doi.org/10.1111/j.1524-4733.2007.00272.x>
- Labiris G, Katsanos A, Fanariotis M et al (2008) Psychometric properties of the Greek version of the NEI-VFQ 25. *BMC Ophthalmol* 8:4. <https://doi.org/10.1186/1471-2415-8-4>
- Cassard SD, Patrick DL, Damiano AM et al (1995) Reproducibility and responsiveness of the VF-14. An index of functional impairment in patients with cataracts. *Arch Ophthalmol* 113:1508–1513. <https://doi.org/10.1001/archophth.1995.01100120038005>
- Hassell JB, Lamoureux EL, Keeffe JE (2006) Impact of age related macular degeneration on quality of life. *Br J Ophthalmol* 90:593–596. <https://doi.org/10.1136/bjo.2005.086595>
- Pondorfer SG, Terheyden JH, Heinemann M et al (2019) Association of vision-related quality of life with visual function in age-related macular degeneration. *Sci Rep* 9:15326. <https://doi.org/10.1038/s41598-019-51769-7>
- Tan JCK, Nguyen V, Fenwick E et al (2019) Vision-related quality of life in keratoconus: a save sight keratoconus registry study. *Cornea* 38:600–604. <https://doi.org/10.1097/ICO.0000000000001899>
- Finger RP, Guymer RH, Gillies MC et al (2014) The impact of anti-vascular endothelial growth factor treatment on quality of life in neovascular age-related macular degeneration. *Ophthalmology* 121:1246–1251. <https://doi.org/10.1016/j.ophtha.2013.12.032>
- Kanellopoulos AJ (2019) The impact of keratoconus treatment with the Athens Protocol (partial topography-guided photorefractive keratectomy combined with higher-fluence corneal collagen cross-linking) on quality of life: a long-term study. *Clin Ophthalmol* 13:795–803. <https://doi.org/10.2147/OPHTH.S188519>
- de Boer MR, Moll AC, de Vet HCW et al (2004) Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt* 24:257–273. <https://doi.org/10.1111/j.1475-1313.2004.00187.x>
- Marakis TP, Koutsandrea C, Poulou MS (2020) The Impact of Vision Impairment on vision-related quality of life of patients with neovascular age-related macular degeneration. *Eur J Ophthalmol* 32:481–490. <https://doi.org/10.1177/1120672120972625>

22. Ratanasukon M, Tongsomboon J, Bhurayanontachai P et al (2016) The Impact of Vision Impairment (IVI) questionnaire; validation of the Thai-version and the implementation on vision-related quality of life in Thai rural community. *PLoS ONE* 11:e0155509. <https://doi.org/10.1371/journal.pone.0155509>
23. Weih LM, Hassell JB, Keeffe J (2002) Assessment of the Impact of Vision Impairment. *Invest Ophthalmol Vis Sci* 43:927–935
24. Evans JR, Mathur A (2005) The value of online surveys. *Internet Res* 15:195–219. <https://doi.org/10.1108/10662240510590360>
25. Frost NA, Sparrow JM, Hopper CD et al (2001) Reliability of the VCM1 questionnaire when administered by post and by telephone. *Ophthalmic Epidemiol* 8:1–11. <https://doi.org/10.1076/opep.8.1.1.1539>
26. Wolffsohn JS, Cochrane AL, Watt NA (2000) Implementation methods for vision related quality of life questionnaires. *Br J Ophthalmol* 84:1035–1040. <https://doi.org/10.1136/bjo.84.9.1035>
27. Mokkink LB, Prinsen CAC, Patrick DL et al. (2019) COSMIN study design checklist for patient-reported outcome measurement instruments. https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf. Accessed 10 Oct 2023
28. Lamoureux EL, Pallant JF, Pesudovs K et al (2006) The Impact of Vision Impairment questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci* 47:4732–4741. <https://doi.org/10.1167/IOVS.06-0220>
29. Keeffe JE, McCarty CA, Hassell JB et al (1999) Description and measurement of handicap caused by vision impairment. *Aust N Z J Ophthalmol* 27:184–186. <https://doi.org/10.1046/j.1440-1606.1999.00179.x>
30. Hassell JB, Weih LM, Keeffe JE (2000) A measure of handicap for low vision rehabilitation: the Impact of Vision Impairment profile. *Clin Exp Ophthalmol* 28:156–161. <https://doi.org/10.1046/j.1442-9071.2000.00312.x>
31. Terheyden JH, Ponderfer SG, Behning C et al (2023) Disease-specific assessment of vision impairment in low luminance in age-related macular degeneration - a MACUSTAR study report. *Br J Ophthalmol* 107:1144–1150. <https://doi.org/10.1136/bjophthalmol-2021-320848>
32. Wright BD, Linacre JM (1989) Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 70:857–860
33. Linacre JM (2006) A user's guide to Winsteps/Ministep: Rasch-model computer programs. Mesa Press, Chicago
34. Waterbury GT (2019) Missing data and the Rasch model: the effects of missing data mechanisms on item parameter estimation. *J Appl Meas* 20:154–166
35. Herzog AR, Rodgers WL (1988) Interviewing older adults: mode comparison using data from a face-to-face survey and a telephone resurvey. *Public Opin Q* 52:84–99
36. Hochstim JR (1967) A critical comparison of three strategies of collecting data from households. *J Am Stat Assoc* 62:976. <https://doi.org/10.2307/2283686>
37. Wright BD, Masters GN (1982) Rating scale analysis: Rasch measurement. Mesa Press, Chicago
38. Boone WJ, Noltemeyer A (2017) Rasch analysis: a primer for school psychology researchers and practitioners. *Cogent Educ* 4:202. <https://doi.org/10.1080/2331186X.2017.1416898>
39. Bond TG, Fox CM (2001) Applying the Rasch model: fundamental measurement in the human sciences, 1st edn. Lawrence Erlbaum Associates Publishers, New York
40. Wright BD, Linacre JM (1994) Reasonable mean-square fit values. *Rasch Measure Transact* 8:370–371
41. Zwick R, Thayer DT, Lewis C (1999) An empirical bayes approach to Mantel-Haenszel DIF Analysis. *J Educ Meas* 36:1–28. <https://doi.org/10.1111/j.1745-3984.1999.tb00543.x>
42. Mohamad Adam Bujang, Hon Yoon Khee, Lee Keng Yee (2022) A step-by-step guide to questionnaire validation research. <https://doi.org/10.5281/zenodo.6801208>
43. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1:30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
44. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15:155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
45. Cicchetti DV, Sparrow SA (1981) Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic* 86:127–137
46. Giavarina D (2015) Understanding Bland Altman analysis. *Biochem Med (Zagreb)* 25:141–151. <https://doi.org/10.11613/BM.2015.015>
47. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
48. Vaz S, Falkmer T, Passmore AE et al (2013) The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS ONE* 8:e73990. <https://doi.org/10.1371/journal.pone.0073990>
49. Deming WE (1943) Statistical adjustment of data. J Wiley & Sons Inc, New York
50. Ciccione L, Dehaene S (2021) Can humans perform mental regression on a graph? Accuracy and bias in the perception of scatterplots. *Cogn Psychol* 128:101406. <https://doi.org/10.1016/j.cogpsych.2021.101406>
51. Nichols KK, Mitchell GL, Zadnik K (2002) Performance and repeatability of the NEI-VFQ-25 in patients with dry eye. *Cornea* 21:578–583. <https://doi.org/10.1097/00003226-200208000-00009>
52. Gwaltney CJ, Shields AL, Shiffman S (2008) Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health* 11:322–333. <https://doi.org/10.1111/j.1524-4733.2007.00231.x>
53. Clayton JA, Eydelman M, Vitale S et al (2013) Web-based versus paper administration of common ophthalmic questionnaires: comparison of subscale scores. *Ophthalmology* 120:2151–2159. <https://doi.org/10.1016/j.ophtha.2013.03.019>
54. Terheyden JH, Mekschat L, Ost RAD et al (2022) Interviewer administration corresponds to self-administration of the vision impairment in low luminance (VILL) questionnaire. *Trans Vis Sci Tech* 11:21. <https://doi.org/10.1167/tvst.11.4.21>
55. Goldstein JE, Fenwick E, Finger RP et al (2018) Calibrating the Impact of Vision Impairment (IVI): creation of a sample-independent visual function measure for patient-centered outcomes research. *Translat Vision Sci Technol* 7:38. <https://doi.org/10.1167/tvst.7.6.38>
56. Lamoureux EL, Ferraro JG, Pallant JF et al (2007) Are standard instruments valid for the assessment of quality of life and symptoms in glaucoma? *Optom Vis Sci* 84:789–796. <https://doi.org/10.1097/OPX.0b013e3181334b83>
57. Taylor DJ, Jones L, Edwards L et al (2021) Patient-reported outcome measures in ophthalmology: too difficult to read? *BMJ Open Ophthalmol* 6:e000693. <https://doi.org/10.1136/bmjophth-2020-000693>
58. Ayton LN, Rizzo JF, Bailey IL et al (2020) Harmonization of outcomes and vision endpoints in vision restoration trials: recommendations from the international HOVER taskforce. *Trans Vis Sci Tech* 9:25. <https://doi.org/10.1167/tvst.9.8.25>
59. Hobart J, Cano S (2009) Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 13:iii, ix–x 1–177. <https://doi.org/10.3310/hta13120>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Jan Henrik Terheyden¹  · Reglind A. D. Ost¹ · Charlotte Behning² · Liza Mekschrat¹ · Gamze Bildik¹ · Maximilian W. M. Wintergerst¹  · Frank G. Holz¹ · Robert P. Finger³ 

✉ Jan Henrik Terheyden
jan.terheyden@ukbonn.de

✉ Robert P. Finger
RobertPatrick.Finger@umm.de

¹ Department of Ophthalmology, University Hospital Bonn,
Ernst-Abbe-Str. 2, 53127 Bonn, Germany

² Department of Medical Biometry, Informatics
and Epidemiology, University Hospital Bonn, Bonn,
Germany

³ Department of Ophthalmology, University Hospital
Mannheim, University of Heidelberg, Mannheim, Germany