



L. Schneidewind<sup>1</sup> · H. Borgmann<sup>2</sup> · S. Schmidt<sup>3</sup>

<sup>1</sup> Institut für Virologie, Universitätsklinikum des Saarlandes, Universität des Saarlandes, Homburg/Saar, Deutschland

<sup>2</sup> Urologische Klinik und Poliklinik, Universitätsmedizin Mainz, Mainz, Deutschland

<sup>3</sup> UroEvidence@DGU, Berlin, Deutschland

# Beantwortung von Selbstbeurteilungsfragebögen mit mobilen Apps im Vergleich zu anderen Methoden

## Originalpublikation

Marcano Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison CP, Car J (2015) Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. *Cochrane Database Syst Rev* (7):MR000042. doi:10.1002/14651858.MR000042.pub2

## Übersetzung

von L. Schneidewind, H. Borgmann und S. Schmidt

## Hintergrund

Selbstbeurteilungsfragebögen sind ein wichtiges Werkzeug zur Datenakquise in der klinischen Praxis, der Gesundheitsforschung und der Epidemiologie. Sie sind ideal, um eine weite geographische Abdeckung der Zielpopulation zu erreichen, sensible Themen zu adressieren und sie sind ressourcenschonender als andere Datenerhebungsmethoden. Diese Selbstbeurteilungsfragebögen können elektronisch zugestellt werden, was die Skalierbarkeit und Geschwindigkeit der Datensammlung bei gleichzeitiger Kostensenkung maximiert. In den letzten Jahren hat die Verwendung von Apps auf Smartphones und Tabletcomputern zu diesem Zweck beträchtliche Aufmerksamkeit erhalten. Jedoch können Unterschiede in der Zustellmethode von

Fragebögen die Qualität der erhaltenen Antworten beeinflussen.

## Ziele

Es soll die Wirkung von Smartphone- und Tablet-Apps als Zustellmethode auf die Qualität der Antworten von Selbstbeurteilungsfragebögen im Vergleich zu alternativen Methoden, wie Papierform, Laptop, Tabletcomputer (hergestellt vor 2007), „short message service“ (SMS) oder plastischen Objekten untersucht werden.

## Suchmethodik

Die Datenbanken MEDLINE, EMBASE, PsycINFO, IEEEExplore, Web of Science, CABI: CAB Abstracts, Current Contents Connect, ACM Digital, ERIC, Sociological Abstracts, Health Management Information Consortium, Campbell Library und CENTRAL wurden durchsucht. ClinicalTrials.gov und die „International Clinical Trials Registry Platform“ der „World Health Organization“ (WHO) wurden weiterhin nach aktuellen und laufenden Studien durchsucht. Zusätzlich wurde nach grauer Literatur in OpenGrey, Mobile Active und ProQuest Dissertation & Theses gesucht. Die Suche wurde schließlich durch „Google Scholar“ und die Referenzlisten der inkludierten Studien und relevanten systematischen Reviews vervollständigt.

Alle Suchen wurden am 12. und 13. April 2015 durchgeführt.

## Auswahlkriterien

Es wurden alle randomisierten kontrollierten Studien, Crossover-Studien und Studien mit gepaart wiederholten Messungen eingeschlossen, die die elektronische Zustellung von Selbstbeurteilungsfragebögen via Smartphone- oder Tablet-App mit einer anderen Zustellungsmethode verglichen. Es wurden alle Daten von Teilnehmern eingeschlossen, die einen gesundheitsbezogenen Selbstbeurteilungsfragebogen (validiert und nicht-validiert) vollständig beantwortet hatten. Außerdem wurden Daten von gesunden Freiwilligen und Teilnehmern mit jeglicher klinischer Diagnose eingeschlossen, sowie Studien, die einen der folgenden Endpunkte berichteten: Datenäquivalenz, Datengenauigkeit, Datenvollständigkeit, Antwortraten, Unterschiede in der benötigten Zeit zur Beantwortung des Selbstbeurteilungsfragebogens, Unterschiede in der Einhaltung des ursprünglichen Stichprobenprotokolls und Annehmbarkeit der Zustellmethode durch die Teilnehmer. Studien, die 2007 oder später publiziert wurden, wurden eingeschlossen, da zu dieser Zeit die Geräte auf den Markt kamen, deren Betriebssysteme App-fokussiert sind.

### Datensammlung und Datenanalyse

Zwei Review-Autoren extrahierten unabhängig voneinander die Daten der eingeschlossenen Studien unter Verwendung eines für dieses Review in REDCap erstellten standardisierten Datensextraktionsformulars.

Die Autoren verglichen dann die Formulare, um einen Konsens zu erreichen. Durch eine initiale systematische Zuordnung der eingeschlossenen Studien wurden zwei Settings identifiziert, in denen die Umfragen durchgeführt wurden: kontrollierte und unkontrollierte. Diese Settings unterschieden sich in folgenden Punkten: (1) Ort zur Beantwortung der Umfrage (2), Frequenz und Intensität des Stichprobenprotokolls (3) und Level der Kontrolle über potenzielle Confounder (z. B. Technologie, Maß an Hilfestellung für die Teilnehmer). Eine narrative Evidenzsynthese wurde erstellt, da eine Metaanalyse aufgrund der hohen klinischen und methodischen Heterogenität ungeeignet war. Die Ergebnisse für jeden Endpunkt wurden entsprechend des Settings der eingeschlossenen Studien berichtet.

### Hauptergebnisse

Es wurden 14 Studien (15 Publikationen) mit insgesamt 2275 Teilnehmern eingeschlossen. Jedoch wurden nur 2272 Teilnehmer in der finalen Analyse berücksichtigt, da die Daten von 3 Studienteilnehmern aus einer Studie fehlten.

Bezüglich der Datenäquivalenz konnte in den inkludierten Studien, sowohl in kontrollierten als auch in unkontrollierten Settings, kein signifikanter Unterschied hinsichtlich der durchschnittlichen Gesamtpunktzahlen zwischen Apps und anderen Zustellmethoden gefunden werden. Auch die Korrelationskoeffizienten überschritten die empfohlenen Schwellenwerte für Datenäquivalenz. Bezüglich der Zeit zur vollständigen Beantwortung eines Selbstbeurteilungsfragebogens zeigte eine Studie in einem kontrollierten Setting, dass die App schneller war als die Beantwortung der Papierversion, während eine andere Studie im gleichen Setting keine si-

gnifikanten Unterschiede zwischen den verschiedenen Arten der Durchführung zeigte. In einem unkontrollierten Setting zeigte eine Studie, dass eine App schneller war als eine SMS.

Die Datenvollständigkeit und Einhaltung des Stichprobenprotokolls wurden nur in unkontrollierten Settings berichtet. Hier erzielte die App mehr vollständig ausgefüllte Fragebögen im Vergleich zur Papierversion und signifikant mehr Dateneingaben als eine SMS. Hinsichtlich der Einhaltung des Stichprobenprotokolls sind Apps wahrscheinlich besser als das Papierformat aber ohne Unterschied zur SMS.

Verschiedene Definitionen zur Akzeptanz der Befragten mit nicht eindeutigen Ergebnissen wurden identifiziert: Präferenz, Einfachheit der Benutzung und Bereitschaft, einen Abgabemodus zu verwenden; Zufriedenheit; Effektivität des Systems zur Bereitstellung von Informationen; wahrgenommene Dauer zur Fertigstellung des Fragebogens; wahrgenommener Nutzen der Art der Durchführung; wahrgenommene Zweckmäßigkeit der Art der Durchführung; wahrgenommene Fähigkeit den Selbstbeurteilungsfragebogen zu beantworten; maximale Dauer, die der Teilnehmer diese Art der Durchführung benutzen würde; Reaktionsfähigkeit der Art der Durchführung und erfolgreiche Einbindung der Art der Durchführung in den Alltag der Teilnehmer.

Schlussendlich berichtete keine der eingeschlossenen Studien, unabhängig vom Setting, Informationen zu Datengenauigkeit und Antwortraten.

### Schlussfolgerung der Autoren

Basierend auf einer narrativen Evidenzsynthese weisen die Ergebnisse darauf hin, dass Apps wahrscheinlich nicht die Datenäquivalenz beeinflussen solange die klinische Anwendung, die Anwendungsfrequenz und das Setting der Validierung nicht verändert werden. Es gab keine Daten zur Datengenauigkeit oder zu Antwortraten. Außerdem waren die Daten zur benötigten Dauer für die Beantwortung eines Selbstbeurteilungsfragebogens widersprüchlich. Obwohl Apps wahrscheinlich die Datenvollständigkeit

verbessern, gibt es nicht genug Evidenz, um ihren Einfluss auf die Einhaltung des Stichprobenprotokolls festzustellen. Keine der eingeschlossenen Studien beurteilte, ob und wie Elemente von Benutzer-Interaktions-Designs, Design des Fragebogens und Interventionsdesign die Methode beeinflussen könnten. Wissenschaftler im Bereich Public Health und Epidemiologie sollten nicht davon ausgehen, dass methodische Effekte, die bei anderen Arten der Durchführung beobachtet wurden, auch für Smartphone- und Tablet-Apps gelten. Wissenschaftler, die in der Methodenentwicklung tätig sind, können die in diesem systematischen Review hervorgehobenen offenen Fragestellungen weiter explorieren.

### Kommentar

von L. Schneidewind

In den letzten Jahren sind auch Apps für Umfragen mit Studienfragebögen interessant geworden. Allerdings herrscht ein Mangel an Daten zur umfassenden Beurteilung von methodischen Effekten bei der Anwendung von Apps zur Datenerhebung. Apps könnten z. B. die Qualität der gesammelten Antworten hinsichtlich Datenäquivalenz, Datengenauigkeit, Datenvollständigkeit und Antwortraten beeinflussen. Im Gegensatz zu dieser methodischen Überlegung steht die hohe Anzahl von Wissenschaftlern, die bereits Apps für Studienfragebögen bei der Datenerhebung verwenden. Aus diesem Grund ist der vorliegende Cochrane Review von großer Bedeutung und sehr aktuell, da zu erwarten ist, dass zukünftig immer mehr Apps insbesondere in der Versorgungs- und epidemiologischen Forschung angewendet werden. Weiterhin werden aus den erhobenen epidemiologischen Daten oft Schlussfolgerungen sowie Entscheidungen für das öffentliche Gesundheitswesen abgeleitet, sodass ein Mindestmaß an methodischer Qualität und Datenäquivalenz unerlässlich ist. Die Aktualität dieses systematischen Reviews aus dem Jahr 2015 steht außer Frage, vor allem auch, weil in einer Updatesuche im August 2015 in MEDLINE keine weiteren Studien identifiziert werden konnten, die die Einschlusskriterien des Reviews erfüllen.

Bereits in der Einleitung dieser Arbeit zeigen die Autoren Limitationen und Schwierigkeiten auf, die ein solcher Review mit sich bringen kann. So wird bei den eingeschlossenen Studien eine Vielzahl von Smartphones, Tabletcomputern und Apps verwendet, was zu einer eingeschränkten Beurteilbarkeit sowie Vergleichbarkeit führt. Außerdem werden verschiedenste Patientenkohorten, z. B. hinsichtlich Alter und Bildung untersucht, was wiederum die Vergleichbarkeit einschränkt. Doch bildet dieser Review einen wichtigen ersten Schritt, um einschätzen zu können, ob die Anwendung von Apps bei Studienfragebögen die Datenqualität beeinflusst. Zusätzlich ist er essenziell für die Methodenforschung und wirft wichtige Fragestellungen für zukünftige Forschungsprojekte auf.

Insgesamt ist dieser Review methodisch qualitativ hochwertig. Positiv hervorzuheben ist v. a. die umfassende Literatursuche, bei der ebenfalls psychologische und methodische Datenbanken berücksichtigt worden sind. Trotz dieser umfassenden Suche konnten lediglich 14 Studien (15 Publikationen) für diesen Review identifiziert werden. Dies scheint im Hinblick auf die derzeit verfügbaren Gesundheits-Apps doch eine geringe Anzahl zu sein. Allerdings ist die Datenlage zur Anwendung und Validierung von Apps in der Gesundheitsforschung sehr begrenzt. Die Autoren warnen explizit davor, Schlussfolgerungen aus anderen Abgabemethoden ungefiltert auf Apps zu übertragen. Ein minimaler negativer Aspekt dieses Reviews ist die Tatsache, dass nur Studien aus Ländern mit hohem Einkommen eingeschlossen werden konnten. Zum einen können so sozioökonomische Faktoren bzw. Barrieren für solche Studienfragebögen nicht identifiziert werden. Zum anderen scheint eine Validierung dieser Methode gerade in wirtschaftlich schwachen Ländern aufgrund der potenziellen Ressourcensparnis gerade sinnvoll. Weiterhin kann diese Methode von großem Nutzen z. B. für sehgeschwache Menschen sein, da Apps auch durch Sprachsteuerung Daten erheben könnten. Daher erscheint eine Evaluation in verschiedensten Studienpopulationen ebenfalls sinnvoll.

Wenn man einen Fragebogen für eine Studie aussucht, sollte bedacht werden, dass die Umstände der eigenen Studie mit denen, in der das jeweilige Instrument validiert worden ist, übereinstimmen. Zusätzlich sollten Forscher bedenken, dass Instrumente speziell für Apps vielleicht anders entwickelt und validiert werden müssen als die konventionelle Papierform, wenn man z. B. an die Länge der verwendeten Fragen denkt.

Insgesamt schlussfolgern die Autoren dieses Cochrane Reviews, dass mit einer App die Datenäquivalenz nicht beeinflusst zu sein scheint, solange die Abgabemethode und klinische Anwendung nicht verändert werden. Aber Forscher sollten auch die Validität und Reliabilität von Studienfragebögen beachten und durch zukünftige Untersuchungen in den neuen Anwendungen besser verstehen lernen. Zwei wichtige Fragestellungen sollten bei der Verwendung von Apps immer Beachtung finden: Zum einen, wie bereits erwähnt: In welchem Setting wurde das entsprechende Instrument validiert und ist das auf meine Untersuchung übertragbar? Zum anderen: Wer ist der Adressat dieses Studienfragebogens und kann er mit der daraus abgeleiteten App umgehen, braucht er Unterstützung bzw. Einweisung in die App?

Widersprüchliche Daten zeigen sich bei der Fragestellung, ob Apps tatsächlich schneller bei der Datenerhebung sind, wobei zu dieser Fragestellung nur zwei Studien aus dem Review tatsächlich Antworten geben. Die Gründe für kontroverse Ergebnisse können sehr unterschiedlich sein, wie der unterschiedliche Aufbau sowie die Schnelligkeit der Apps oder die Einweisung vor dem Gebrauch der App.

Weiterhin konnte dieser Review keine Überlegenheit von elektronischen Datenerhebungsmethoden bezüglich der Einhaltung des Protokolls zeigen. Dies steht im Kontrast zu zwei anderen publizierten Reviews, in denen die Autoren zur der Schlussfolgerung kamen, dass elektronische Methoden überlegen sind [1, 2].

Ein wichtiger Kritikpunkt der Autoren an den eingeschlossenen Studien ist, dass in keiner der Publikationen die Mensch-Computer-Interaktion diskutiert wird. Es ist bekannt, dass diese

Interaktion einen wichtigen Aspekt bei der Anwendung von App-basierten Studienfragebögen darstellt und Antworten z. B. durch die Einfachheit des Aufbaus der App, Bedienerfreundlichkeit oder Gestaltung sowie Farbauswahl beeinflusst wird. Außerdem bedeutet die genauere Untersuchung dieser Mensch-Computer-Interaktion auch eine Chance, weil darüber Aspekte identifiziert werden könnten, die die Motivation zur Teilnahme an Umfragen erhöhen.

Für die weitere Forschung und Evaluation von Apps als Abgabemethode von Studienfragebögen sind Kosten-Nutzen-Analysen unbedingt erforderlich. Dieser Fakt wurde in dem vorliegenden Review nicht eruiert. Zusätzlich ist diese Analyse nicht trivial, da es multiple Faktoren bzw. Fragestellungen zu berücksichtigen gilt: Können durch die Anwendung von Apps tatsächlich Kosten gespart werden, insbesondere hinsichtlich Geräewartung, Softwareupdates und Kosten von Schulungspersonal etc.? Gibt es evtl. altersunterschiedliche Populationen bezüglich der Kosten-Nutzen-Relation, z. B. dass ältere Menschen vielleicht schneller eine Papierversion ausfüllen? Hinsichtlich der Ressourcen und unserer Umwelt, was ist besser: App oder Papier?

Alle erwähnten Faktoren verdeutlichen, dass bei der Anwendung von Apps als Abgabemethode von Studienfragebögen noch viel Forschungsarbeit geleistet werden muss und es insgesamt wenig Evidenz zu der Frage gibt, ob Apps einen Vorteil gegenüber herkömmlichen Methoden zur Datenerhebung haben. Daher betonen die Autoren des vorliegenden Reviews wiederholt, dass Methoden, wie sie für papierbasierte Fragebögen etabliert worden sind, nicht unreflektiert übertragen werden können.

Doch wie sieht es in der Urologie aus? In diesem Cochrane Review wurde eine urologische Studie eingeschlossen, die die Validität und Reliabilität des „International Prostate Symptom Scores“ (IPSS) als Smartphone-App untersucht [3]. Hier schlussfolgerten die Autoren, dass die Validität und Reliabilität der App vergleichbar mit der konventionellen Papierversion des IPSS sind. Die Autoren betonen ausdrücklich die Wichtigkeit, Apps vor ihrer Anwendung zu

validieren und diese Validierung in einem vergleichbaren Setting vorzunehmen, in dem die konventionelle Version eines Studienfragebogens geprüft wurde [4].

Trotz all dieser Schwierigkeiten und offenen Probleme bedeuten App-Anwendungen in der Datenerhebung auch eine Chance, und das auch in der Urologie. Heutzutage besitzt fast jeder ein Smartphone, weltweit sollen es mittlerweile ca. 6 Mrd. Smartphones sein [5]. Daher kann eine breite Masse an Patienten erreicht werden. So können z. B. auch Patienten berücksichtigt werden, die in ländlichen Gebieten mit schlechter fachärztlicher Versorgung leben. Auch können Apps in der Versorgungsforschung genutzt werden. Dies könnte auch in der Urologie sinnvoll sein, z. B. zum Monitoring von BPH-Patienten mittels IPSS. Ähnliche Projekte zur Patientenbetreuung in ländlichen Gebieten gibt es bereits in anderen Fachgebieten. So gibt es ein DFG-gefördertes psychiatrisches Projekt zur Betreuung von alkoholabhängigen Patienten in Mecklenburg-Vorpommern [6].

Daher sollten auch wir Urologen diese Chance nutzen, Apps als Abgabemethode für Studienfragebögen zu etablieren. Damit dies sinnvoll geschehen kann, ist vor allem methodische Forschung und die Validierung von Apps für bereits anerkannte Fragebögen notwendig. Nur so wird man Informationen darüber erhalten, ob mit weniger Ressourcen mehr verlässliche gesundheitsbezogene (Langzeit-)Daten erhoben werden können mit dem Ziel, die Versorgung unserer Patienten stetig zu verbessern.

### Fazit für die Praxis

- Die momentan verfügbare Evidenz reicht nicht aus, um klare Empfehlungen für die Anwendung von Apps auszusprechen.
- Die Datenäquivalenz scheint bei einer App als Abgabemethode von Studienfragebögen im Vergleich zu anderen Abgabemethoden nicht beeinflusst zu sein, solange das Patientenkollektiv im Vergleich zum Setting der ursprünglichen Frage-

**bogenvvalidierung nicht verändert wird.**

- **Bezüglich Schnelligkeit und Einhaltung des Protokolls sind die Daten im Vergleich zwischen App-basierten und konventionellen Fragebögen widersprüchlich.**
- **Apps bieten mögliche Chancen im Hinblick auf ihre breiten Anwendungsmöglichkeiten sowie ihr Potenzial zu einer großen geographischen Abdeckung und Ressourcenersparnis.**
- **Weitere methodische Forschung und Kosten-Nutzen-Analysen sind notwendig.**
- **Zentrale Themen wie Datensicherheit und Datenschutz müssen adressiert werden.**

### Korrespondenzadresse

#### L. Schneidewind

Institut für Virologie, Universitätsklinikum des Saarlandes, Universität des Saarlandes  
Kirrberger Str. 100, 66424 Homburg/Saar, Deutschland  
laila.schneidewind@uks.eu

**Danksagung.** UroEvidence dankt Ingrid Töws und Katharina Kunzweiler (beide Cochrane Deutschland) für die Durchsicht und Kommentare zu dieser Übersetzung.

### Einhaltung ethischer Richtlinien

**Interessenkonflikt.** L. Schneidewind, H. Borgmann und S. Schmidt geben an, dass kein Interessenkonflikt besteht.

Dieser Beitrag beinhaltet keine von den Autoren durchgeführten Studien an Menschen oder Tieren.

### Literatur

1. Lane SJ, Heddle NM, Arnold E, Walker I (2006) A review of randomized controlled trials comparing effectiveness of hand held computers with paper methods for data collection. BMC Med Inform Decis Mak. doi:10.1186/1472-6947-6-23
2. Gwaltney CJ, Shield 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
3. Kim JH, Kwon SS, Shim SR, Sun HY, Ko YM, Chun DI et al (2014) Validation and reliability of a smartphone application for the International Prostate Symptom Score questionnaire: A randomized repeated measures cross over study. J Med Internet Res 16(2):e38. doi:10.2196/jmir.3042

1. Yaxley JW, Coughlin GD, Chambers SK, Occhipinti S, Samaratunga H, Zajdlwicz L, Dungleison N, Carter R, Williams S, Payton DJ, Perry-Keene J, Lavin MF, Gardiner RA. Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled phase 3 study. Lancet. 2016 Jul 26. pii: S0140-6736(16)30592-X.

Schwerpunkt: Oncology – Urology

- Relevanz für die Praxis: ■ ■ ■ ■ ■ ■ ■ ■
- Ist das neu?: ■ ■ ■ ■ ■ □ □

2. Lu Z, Lin G, Reed-Maldonado A, Wang C, Lee YC, Lue TF. Low-intensity Extracorporeal Shock Wave Treatment Improves Erectile Function: A Systematic Review and Meta-analysis. Eur Urol. 2016 Jun 16. pii: S0302-2838(16)30259-7.

Schwerpunkt: Oncology – Urology

- Relevanz für die Praxis: ■ ■ ■ ■ ■ □ □
- Ist das neu?: ■ ■ ■ ■ ■ □ □

Empfohlen von: EvidenceUpdates (BMJ und McMaster University)

- 
4. Barry MJ, Fowler FJ Jr, O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK, Cockett AT (1992) The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *J Urol* 5:1549–1557
  5. Nielsen (2010) Smartphones to overtake feature phones in US by 2011. <http://www.nielsen.com/us/en/newswire/2010/smartphones-to-overtake-feature-phones-in-u-s-by-2011.htm>. Zugegriffen: 11.08.2016
  6. Kontinuierliche Behandlung von Patienten mit Alkoholabhängigkeit mit SMS Continuity of care among alcohol dependent patients via mobile phone SMS (CAPS); PI: Michael Lucht; Deutsche Forschungsgemeinschaft (DFG): <http://gepris.dfg.de/gepris/person/1759948>; <http://www.terra-digitalis.dfg.de/27-heilsame-sms.html> <http://www.isrctn.com/ISRCTN78350716>. Zugegriffen: 11.08.2016