Stadt Wien | Smart City | DIGITAL CITY .WIEN

FFG Forschung wirkt. | Bundesministerium Verkehr, Innovation und Technologie

WAALTeR

Aktives Altern findet Stadt

D5.3 Evaluierungsbericht

Datum: 26.02.2020

AutorInnen: Till Bieg (AIT), Isabel Schwaninger (TU), Cornelia Gerdenitsch (AIT), Markus Garschall (AIT), Christopher Frauenberger (TU)

1 Inhalt

2 Zusammenfassung 3
3 Zielsetzung und Fragestellungen der Evaluierung 4
4 Methodisches Vorgehen 6
4.1 Stichprobenziehung und Zuteilung zu den Untersuchungsgruppen 6
4.2 Vorzeitige Beendigung der Studienteilnahme und Rücklauf 9
4.3 Stichprobenbeschreibung 10
5 Ergebnisse 15
5.1 Studienphase 1 15
5.2 Studienphase 2 34
5.3 Ergebnisse aus den qualitativen Interviews 38
5.4 Weitere Analysen 40
5.5 Zusammenfassung der Ergebnisse in Bezug auf die Fragestellungen 42
6 Diskussion 44
6.1 Reflexion und Empfehlungen für zukünftige Studien 44
6.2 Konklusion 46
7 Literaturverzeichnis 47

2 Zusammenfassung

Dieses Deliverable beschreibt die Ergebnisse der WAALTeR-Evaluierungsstudie, die von März 2018 bis September 2019 in Wien durchgeführt wurde. Der genaue zeitliche Ablauf der Evaluierungsstudie und methodische Überlegungen sind in Deliverable 5.1 – Evaluierungsplan (D5.1) beschrieben. Ergänzend zu D5.1 fasst das vorliegende Deliverable den theoretischen Hintergrund und zugrundeliegende Fragestellungen zusammen und beschreibt, wenn erforderlich, weitere methodische Aspekte der Evaluierungsstudie.

Zur Evaluierung des WAALTeR-Systems wurde eine 18-monatige Längsschnittstudie mit mehr als 140 TeilnehmerInnen durchgeführt. Dabei wurden Nutzungsdaten sowie Daten basierend auf quantitativen und qualitativen Selbstberichten erfasst. Ein Schwerpunkt der Evaluierung lag auf der möglichen Wirkung von Technologie auf Selbstbestimmtheit, Autonomie, Soziale Partizipation und Sicherheitsempfinden von älteren Menschen. StudienteilnehmerInnen wurden zufällig zu einer Testgruppe (mit WAALTeR-System) oder einer Vergleichsgruppe (ohne WAALTeR-System) zugeteilt. In einer zweiten Studienphase hatten TeilnehmerInnen der Testgruppe darüber hinaus die Möglichkeit zusätzliche Technologien auszuprobieren. Die im Rahmen der Studie erhobenen Daten sprechen nicht für signifikante Veränderungen in den untersuchten Messgrößen innerhalb der Untersuchungsgruppen beziehungsweise über den Zeitraum der Studie. Nutzungsdaten, Daten zur Bewertung des WAALTeR-Systems und Erkenntnisse aus qualitativen Interviews ergänzen die Analyse und zeigen mögliche Gründe für dieses wichtige Ergebnis auf. Sie geben darüber hinaus wertvolle Einblicke, welche Bedeutung digitale Technologien im Leben älterer Menschen haben können. Abschließend werden methodische Limitationen und Implikationen für zukünftige Untersuchungen diskutiert.

3 Zielsetzung und Fragestellungen der Evaluierung

Durch den demografischen Wandel ergibt sich in vielerlei Hinsicht das Interesse, neue Wege zu finden, um älteren Menschen ein aktives, selbstbestimmtes Leben zu ermöglichen. Insbesondere neuen Technologien aus dem Bereich „Active and Assisted Living“ (AAL) besitzen dabei das Potential, ältere Menschen in ihrem Alltag zu unterstützen und so zu einer selbstbestimmten Lebensführung beizutragen. Dadurch könnten AAL-Technologien einen Beitrag leisten, um die Herausforderungen des demografischen Wandels zu bewältigen.

Vor diesem Hintergrund entwickelte sich in den vergangenen Jahren auch ein Interesse daran, den potentiellen Mehrwert solcher Technologien wissenschaftlich zu evaluieren. Eine aktuelle Studie von Dupuy, Conel und Sauzéon (2016) untersuchte die Wirkung einer tabletbasierten Technologie auf die Selbstbestimmtheit älterer Menschen in einem quasiexperimentellen Design. Bemerkenswerterweise berichteten Personen der Versuchsgruppe nach der sechsmonatigen Intervention im Vergleich zur Kontrollgruppe eine signifikant höhere Selbstbestimmtheit.

Insgesamt ist empirische Evidenz, die für eine positive Wirkung von Technologie auf Facetten der Lebensqualität älterer Menschen spricht, allerdings kaum vorhanden (Peek, Aarts, & Wouters, 2015). Bisherigen Studien fehlt häufig die notwendige methodische Strenge, um fundierte Aussagen treffen zu können (Morris et al., 2013; Reeder et al., 2013): Oft handelt es sich um Fallstudien und die Menge an experimentellen Untersuchungen – insbesondere mit größeren Stichproben – ist gering.

Ein wichtiges Anliegen im Rahmen der WAALTeR-Evaluierungsstudie war daher die Umsetzung eines Studiendesigns, das die Untersuchung von kausalen Zusammenhängen erlaubt. Entsprechend wurde ein experimentelles Design mit mehreren Messzeitpunkten umgesetzt. Die Auswahl abhängiger Variablen erfolgte dabei einerseits auf der Grundlage bisheriger Forschungsergebnisse und andererseits unter Berücksichtigung der spezifischen Funktionalität des WAALTeR-Systems. Es wurden jene Konstrukte ausgewählt, auf die die Nutzung des WAALTeR-Systems plausiblerweise einen positiven Einfluss haben könnte. Beispielsweise lässt sich annehmen, dass die Nutzung einer Smartwatch mit Notruf-Funktion das subjektive Sicherheitsempfinden älterer Menschen erhöht. Ebenso erscheint es möglich, dass sich die Bereitstellung eines digitalen Nachbarschaftsnetzwerks auf die soziale Partizipation älterer Menschen auswirkt. Zur Operationalisierung der Konstrukte wurden nach Möglichkeit normierte, psychometrisch validierte Skalen verwendet.

Allerdings beschränkt sich die Studie nicht nur auf ein quantitatives Untersuchungsdesign, das Veränderungen in verschiedenen abhängigen Variablen untersucht: Darüber hinaus werden Daten zur Nutzung und Bewertung der Technologie mittels quantitativer und qualitativer Erhebungsmethoden erfasst. Diese Daten können dabei helfen, Ergebnisse aus der Wirkungsanalyse besser zu verstehen und zu interpretieren. Sie ermöglichen eine ganzheitliche Betrachtung individueller Bedürfnisse und Rahmenbedingungen zur Nutzung der Technologie.

Auf Grundlage dieser Überlegungen werden im Rahmen dieses Deliverables folgende Fragestellungen untersucht:

- Fragestellung 1: Beeinflusst die Nutzung des WAALTeR-Systems die Selbstbestimmtheit, Autonomie, Soziale Partizipation und das Sicherheitsempfinden älterer Menschen?
- Fragestellung 2: Wie wurde das WAALTeR-System genutzt (Häufigkeit, Dauer, Anwendungen)?
- Fragestellung 3: Wie wird das WAALTeR-System hinsichtlich seiner User Experience von den NutzerInnen bewertet?
- Fragestellung 4: Welche Rolle haben Technologien wie WAALTeR in der Lebenswelt älterer Menschen?

4 Methodisches Vorgehen

In Ergänzung zu D5.1 beschreibt der folgende Abschnitt weitere methodische Details der Evaluierungsstudie. Grundsätzlich gliedert sich die WAALTeR-Evaluierungsstudie in zwei Studienphasen. In Studienphase 1 wurden Personen der Testgruppe mit Tablet und Uhr ausgestattet. Diese Studienphase dauerte zwölf Monate. In Studienphase 2 hatten die TeilnehmerInnen der Testgruppe zusätzlich die Möglichkeit weitere Technologie nach Bedarf und Interesse auszuprobieren. Diese Studienphase dauerte sechs Monate.

4.1 Stichprobenziehung und Zuteilung zu den Untersuchungsgruppen

4.1.1 Studienphase 1

Die untersuchte Stichprobe wurden im Allgemeinen über verschiedene Medienkanäle (Anzeigen im Internet, Printmedien etc.) rekrutiert. An einer Teilnahme interessierte Personen wurden im Rahmen von Informationsveranstaltungen über die Studie und Teilnahmebedingungen informiert und gaben in Folge ihr Einverständnis zur Studienteilnahme. Somit handelt es sich bei den untersuchten Personen um eine Gelegenheitsstichprobe, zu deren Auswahl keine probabilistischen Methoden angewendet wurden. Dieses Vorgehen wurde aufgrund der praktischen Herausforderungen bei der Rekrutierung (ausreichende Stichprobengröße; keine Möglichkeit zum Rückgriff auf Datensätze, die die Anwendung probabilistischer Sampling-Methoden erlauben) gewählt.

Rekrutierte StudienteilnehmerInnen wurde vor Beginn der ersten Studienphase randomisiert zu Test- oder Vergleichsgruppe zugeteilt. Dies stellt einen zentralen Aspekt des experimentellen Designs der Studie dar. Generell erlaubt die randomisierte (das heißt zufallsbasierte) Zuteilung von TeilnehmerInnen zu verschiedenen Versuchsbedingungen die Untersuchung kausaler Effekte und gilt als methodischer Gold-Standard für das Design von Interventionsstudien (Bortz & Döring, 2006). Im Kontext der WAALTeR-Evaluierungsstudie erlaubt dieses Untersuchungsdesign die Wirkung der Nutzung des WAALTeR-Systems auf ausgewählte Outcomes (Selbstbestimmtheit, Sicherheitsempfinden, Soziale Partizipation, etc.) quantitativ zu untersuchen.

Aufgrund des hohen Interesses an der Studienteilnahme durch Ehepaare wurde entschieden, dass Personen, die gemeinsam in einem Haushalt leben, auch an der Studie teilnehmen können. Um die Unabhängigkeit der Datenstruktur zu gewährleisten, wurden bei TeilnehmerInnen, die gemeinsam in einem Haushalt leben (n = 8), jeweils eine Person per Zufall von der weiteren Analyse ausgeschlossen (n = 4). Die jeweilige Person wurde *vor* der randomisierten Zuteilung zu den Untersuchungsgruppen zufällig ausgesucht. Dieses Vorgehen wurde aufgrund der schlecht balancierten hierarchischen Datenstruktur gewählt, bei der nur verhältnismäßig wenige TeilnehmerInnen gemeinsam in einem Haushalt leben, was die adäquate Abbildung der Datenstruktur, beispielsweise mittels hierarchischer Modellen, erschwert. Entsprechend werden alle weiteren Analysen, die die Unabhängigkeit der Datenstruktur voraussetzen, mit diesem reduzierten Datensatz durchgeführt. Das bedeutet, dass pro Haushalt jeweils nur eine zufällig ausgewählte Person in die jeweilige Analyse miteinbezogen wurde.

Die TeilnehmerInnen, die im Rahmen der Auftaktveranstaltungen ihr Einverständnis zur Studienteilnahme gegeben haben (*n* = 167), wurden zu einer der folgenden Untersuchungsgruppen zugeteilt:

- Testgruppe *mit* WAALTeR-System (Tablet und Uhr), erhalten Angebote zur freiwilligen Teilnahme an Stammtischen, Vorträgen, Workshops etc.
- Vergleichsgruppe *ohne* WAALTeR-System, erhalten Angebote zur freiwilligen Teilnahme an Stammtischen, Vorträgen, Workshops etc.

Dabei wurde eine randomisierte, proportional stratifizierte Zuteilungsmethode gewählt. Dieser Ansatz wirkt Verzerrungen in der Verteilung von Variablen innerhalb beider Untersuchungsgruppen entgegen, die trotz einer randomisierten Zuteilung entstehen könnten. Proportional stratifiziert bedeutet im Fall der WAALTeR-Evaluierungsstudie, dass relevante Merkmale (Merkmale, die für die Zusammensetzung der Stichprobe und zur Untersuchung der wissenschaftlichen Fragestellungen von zentraler Bedeutung sind) in beiden Untersuchungsgruppen annährend gleich verteilt sind wie in der Gesamtstichprobe. Dabei wurde die Stichprobe nach folgenden Merkmalen stratifiziert:

- Geschlecht
- Alter in Jahren
- Self Determination Satisfaction; gemessen mit Balanced Measure of Psychological Needs (BMPN; Neubauer & Voss, 2016)

Die Anzahl an Stratifizierungsvariablen ist begrenzt durch die Stichprobengröße bzw. die Anzahl an Personen in den Subgruppen, die auf Basis der Stratifizierungsvariablen gebildet werden. Die aus der Zuteilung resultierenden Stichprobenkennwerte dieser Merkmale in beiden Untersuchungsgruppen und der Gesamtstichprobe sind Tabelle 1 bis 3 dargestellt.[1] Zusammenfassend zeigen sich dabei nur minimale Unterschiede, was für die Vergleichbarkeit der Gruppen spricht. Diese Vergleichbarkeit bildet die methodische Grundlage zur Untersuchung kausaler Effekte innerhalb der WAALTeR-Evaluierungsstudie.

Tabelle 1. Stichprobekennwerte der Variable Alter (in Jahren) in der Gesamtstichprobe und den zwei Untersuchungsgruppe (Testgruppe und Vergleichsgruppe) vor dem ersten Messzeitpunkt (1.0).

Gruppe	*M (SD)*	*Mdn*	*Min*	*Max*
Testgruppe (*n* = 91)	72.78 (6.48)	73.00	61	89
Vergleichsgruppe (*n* = 76)	72.01 (6.59)	71.00	60	88
Gesamtstichprobe (*n* = 167)	72.43 (6.52)	72.00	60	89

Anmerkungen. M = Mittelwert. *SD* = Standardabweichung. *Mdn* = Median. *Min* = Minimum. *Max* = Maximum.

[1]Im Rahmen dieses Berichts wird bei der Darstellung von Zahlen als Dezimaltrennzeichen das Punktsymbol (**.**) verwendet.

Tabelle 2. Stichprobekennwerte der Variable Self Determination Satisfaction in der Gesamtstichprobe und den zwei Untersuchungsgruppen (Testgruppe und Vergleichsgruppe) vor dem ersten Messzeitpunkt (1.0).

Gruppe	*M (SD)*	*Mdn*
Testgruppe (*n* = 91)	5.83 (0.83)	6.00
Vergleichsgruppe (*n* = 76)	5.83 (0.76)	5.89
Gesamtstichprobe (*n* = 167)	5.83 (0.79)	5.89

Anmerkungen. M = Mittelwert. *SD* = Standardabweichung. *Mdn* = Median.

Tabelle 3. Relative Häufigkeiten der Variable Geschlecht in der Gesamtstichprobe und den zwei Untersuchungsgruppen (Testgruppe und Vergleichsgruppe) vor dem ersten Messzeitpunkt (1.0).

	Relative Häufigkeit (in %)	
Gruppe	Frauen	Männer
Testgruppe (*n* = 91)	72.53	27.47
Vergleichsgruppe (*n* = 76)	75.00	25.00
Gesamtstichprobe (n = 167)	73.65	26.35

4.1.2 Studienphase 2

Im Rahmen der zweiten Studienphase wurden TeilnehmerInnen aus der Testgruppe (TeilnehmerInnen mit Tablet und Uhr) mit zusätzlichen Technologien ausgestattet (siehe auch D5.1). Die Zuteilung erfolgte dabei nicht randomisiert, sondern ausschließlich nach Bedarf und Interesse. Dabei konnten die TeilnehmerInnen zwischen einer der folgenden Optionen wählen[2]:

- Keine zusätzliche Technologie – Nutzung des WAALTeR-Systems (Tablet und Uhr) wie bisher
- Telemedizin-Paket zusätzlich zu Tablet und Uhr
- Sicherheitspaket zusätzlich zu Tablet und Uhr

Entsprechend wurden im Rahmen der zweiten Studienphase insgesamt vier Untersuchungsgruppen betrachtet (Gesamtstichprobe *n* = 144):

- Testgruppe ohne zusätzliche Technologie (*n* = 42)
- Testgruppe mit Telemedizin-Paket (*n* = 31)
- Testgruppe mit Sicherheitspaket (*n* = 11)
- Vergleichsgruppe (*n* = 60)

[2] Für eine ausführliche Beschreibung der zusätzlichen Technologien siehe D3.4 und D3.5.

Die genaue Zusammensetzung der Stichproben in den verschiedenen Studienphasen sowie die Stichprobe in Bezug auf die qualitativen Interviews sind in den Abschnitten 4.3.1, 4.3.2 und 5.3.1 beschrieben.

4.2 Vorzeitige Beendigung der Studienteilnahme und Rücklauf

4.2.1 Studienphase 1

Von den 167 StudienteilnehmerInnen (91 Personen in Testgruppe, 76 Personen in Vergleichsgruppe) beendeten innerhalb der ersten Studienphase insgesamt 23 Personen ihre Teilnahme vorzeitig (7 Personen in Testgruppe, 16 Personen in Vergleichsgruppe) Eine Person aus der Vergleichsgruppe wurde vor Studienbeginn durch Zufallsauswahl in die Testgruppe nachbesetzt, nachdem eine andere Person, die der Testgruppe zugeteilt war, ihre Teilnahme vorzeitig beendete. Gründe für einen Ausstieg waren z.B. die Zuordnung zur Vergleichs- statt Testgruppe, Umzug und Krankheit.

Alle Rücklaufstatistiken beziehen nur jene Personen, die ihre Studienteilnahme nicht vorzeitig beendet haben. Insgesamt wurden in der ersten Studienphase 87.92% der ausgesendeten Fragebögen retourniert. Tabelle 4 stellt den Rücklauf (Anteil zurückgesendeter Fragebögen im Verhältnis zu allen ausgesendeten Fragebögen) zu den fünf Messzeitpunkten und pro Untersuchungsgruppe dar.

Tabelle 4. Rücklauf (prozentualer Anteil zurückgesendeter Fragebögen im Verhältnis zu allen ausgesendeten Fragebögen) während der ersten Studienphase.

	Rücklauf (in %)		
Messzeitpunkt	Testgruppe	Vergleichsgruppe	Gesamtstichprobe
1.0	95.24	85.00	90.97
1.1	95.24	85.00	90.97
1.2	94.05	81.67	88.89
1.3	92.86	75.00	85.42
1.4 / 2.0	89.29	75.00	83.33
Gesamt	93.33	80.33	87.92

Unabhängig davon, ob sie vorzeitig aus der Studie ausgestiegen sind oder nicht, retournierten insgesamt 17 TeilnehmerInnen überhaupt keinen Fragebogen während der ersten Studienphase. Diese Personen wurden als „Non-Responder“ klassifiziert und von weiteren Analysen ausgeschlossen. Entsprechend beziehen sich Auswertungen in Bezug auf die Studienphase 1, wenn nicht anders bezeichnet, auf die verbleibende Stichprobe mit 150 Personen (89 Personen in Testgruppe, 61 Personen in Vergleichsgruppe).

4.2.2 Studienphase 2

Im Laufe der zweiten Studienphase beendeten 3 weitere Personen (2 Person aus der Testgruppe, 1 Person aus der Vergleichsgruppe) ihre Teilnahme vorzeitig. Somit stiegen insgesamt[3] 26 Personen (9 Person aus der Testgruppe, 17 Person aus der Vergleichsgruppe) aus der Studie aus, was einer Ausstiegsrate von insgesamt 15.57% entspricht.

Tabelle 5. Rücklauf (prozentualer Anteil zurückgesendeter Fragebögen im Verhältnis zu allen ausgesendeten Fragebögen) während der zweiten Studienphase.

	Rücklauf (in %)			
Messzeitpunkt	Kein Zusatz	Telemedizin	Sicherheit	Vergleichs-gruppe
2.1	80.00	87.10	90.90	69.49
2.2	90.00	83.90	90.90	77.97
Gesamt	85.00	85.48	90.90	73.72

Tabelle 5 stellt den Rücklauf (Anteil zurückgesendeter Fragebögen im Verhältnis zu allen ausgesendeten Fragebögen) zu den fünf Messzeitpunkten und pro Untersuchungsgruppe dar. Dabei sind im Allgemeinen niedrigere Rücklaufraten im Vergleich zur ersten Studienphase auffällig. Über beide Studienphasen betrachtet entspricht die Rücklaufrate 87.6%.

4.3 Stichprobenbeschreibung

4.3.1 Studienphase 1

Im Folgenden werden wesentliche demografische Merkmale der in Studienphase 1 untersuchten Stichprobe beschrieben. Die Stichprobe (n = 150) besteht aus Personen wohnhaft in Wien im Alter von 60 Jahren oder älter. Alle Personen gaben an, ihre Wohnung selbstständig verlassen zu können und keine diagnostizierte psychische Erkrankung zu haben (für eine Erläuterung der Teilnahmevoraussetzungen siehe D5.1). Stichprobenkennwerte zu den Variablen Alter, Haushaltgröße, Geschlecht und Bildungstand sind respektive in den Tabelle 6 bis 9 dargestellt. In beiden Untersuchungsgruppen ist die Altersverteilung ähnlich mit einem Mittelwert von ca. 72 Jahren. In Bezug auf die Geschlechterverteilung ist der Anteil an Frauen vergleichsweise hoch: Ungefähr Dreiviertel aller TeilnehmerInnen sind weiblich. Ein Großteil der TeilnehmerInnen lebt alleine oder zu zweit, nur 3% Prozent der TeilnehmerInnen leben in Haushalten mit drei Personen oder mehr. Bezüglich des Bildungsstands (entspricht der höchsten abgeschlossenen Ausbildung) zeigt sich, dass ein Großteil der TeilnehmerInnen eine Fachschule oder Lehre abgeschlossen hat, ungefähr ein Drittel haben die Matura. 17.33 % der TeilnehmerInnen haben einen Universitätsabschluss, am wenigsten vertreten sind Personen, die nur die Pflichtschule abgeschlossen haben (6.67 %).

[3] über beide Studienphasen betrachtet

Tabelle 6. Stichprobekennwerte der Variable *Alter* (in Jahren) in der untersuchten Stichprobe zu Studienphase 1.

Gruppe	*M (SD)*	*Mdn*	*Min*	*Max*
Testgruppe (*n* = 89)	72.78 (6.48)	73.00	61	89
Vergleichsgruppe (*n* = 61)	72.01 (6.59)	71.00	60	88
Gesamt (*n* = 150)	72.43 (6.52)	72.00	60	89

Anmerkungen. M = Mittelwert. *SD* = Standardabweichung. *Mdn* = Median. *Min* = Minimum. *Max* = Maximum.

Tabelle 7. Relative Häufigkeiten der Variable Haushaltsgröße (entspricht der Anzahl von Personen in einem Haushalt) in der untersuchten Stichprobe in Studienphase 1.

	Relative Häufigkeit (in %)		
Gruppe	1 Person	2 Personen	>= 3 Personen
Testgruppe (*n* = 89)	52.28	43.19	4.53
Vergleichsgruppe (*n* = 61)	55.74	42.62	1.64
Gesamt (*n* = 150)	53.91	42.96	3.13

Tabelle 8. Relative Häufigkeiten der Variable Geschlecht der untersuchten Stichprobe in Studienphase 1.

	Relative Häufigkeit (in %)	
Gruppe	Frauen	Männer
Testgruppe (*n* = 89)	71.90	28.09
Vergleichsgruppe (*n* = 61)	77.05	22.95
Gesamt (*n* = 150)	74.00	26.00

Tabelle 9. Relative Häufigkeiten der Variable Bildungsstand (entspricht der höchsten abgeschlossenen Ausbildung) in der untersuchten Stichprobe in Studienphase 1.

	Relative Häufigkeit (in %)			
Gruppe	Pflicht-schulab-schluss	Lehr- / Fachschul-abschluss	Matura	Universitäts-abschluss
Testgruppe (*n* = 89)	7.86	39.33	34.83	17.98
Vergleichsgruppe (*n* = 61)	4.92	52.46	26.23	16.39
Gesamt (*n* = 150)	6.67	44.67	31.33	17.33

Darüber hinaus stellt Tabelle 10 relative Häufigkeiten in Bezug auf den Besitz von Smartphone und Tablet *vor Studienbeginn* dar. Dies kann als (schwacher) Indikator für die Vorerfahrungen mit digitalen Technologien angesehen werden. Diesbezüglich sind durchaus Unterschiede zwischen den Gruppen zu erkennen: Einerseits ist in der Vergleichsgruppe der Besitz von Smartphones geringer als in der Testgruppe. Auf der anderen Seite ist in der Testgruppe der Besitz von Tablets geringer in Referenz zum Prozentsatz an Personen in der Vergleichsgruppe, die ein Tablet besitzen. Insgesamt besaßen vor Studienbeginn bereits ca. Dreiviertel aller TeilnehmerInnen ein Smartphone und ca. ein Drittel ein Tablet, was auf eine relativ technologieerfahrene Stichprobe hinweist.

Insgesamt erscheinen die verschiedenen demografischen Variablen innerhalb der beiden Untersuchungsgruppen ähnlich verteilt. Dies impliziert eine gute Vergleichbarkeit beider Gruppen.

Tabelle 10. Relative Häufigkeiten in Bezug auf den Besitz von Smartphone und Tablet vor Studienbeginn in der untersuchten Stichprobe in Studienphase 1.

	Relative Häufigkeit (in %)	
Gruppe	Personen mit Smartphone	Personen mit Tablet
Testgruppe (*n* = 89)	79.54	28.40
Vergleichsgruppe (*n* = 61)	66.67	35.00
Gesamt (*n* = 150)	74.32	31.08

4.3.2 Studienphase 2

Im Folgenden werden Stichprobekennwerte aufgegliedert nach den Untersuchungsgruppen in Studienphase 2 dargestellt. Wie auch in Studienphase 1 wurden Personen, die nicht mindestens einen Fragebogen zurückgesendet haben, in der weiteren Analyse nicht berücksichtigt (unabhängig davon, ob sie die Studienteilnahme vorzeitig beendet haben oder nicht). Somit umfasst die Gesamtstichprobe in der zweiten Studienphase 140 Personen. Tabelle 11 bis Tabelle 14 stellen die demografischen Variablen Alter, Geschlecht, Bildungsstand und Haushaltsgröße aufgegliedert nach den vier Untersuchungsgruppen dar.

Interessanterweise zeigt sich in der Gruppe von Personen, die das Sicherheitspaket gewählt haben, ein etwas höherer Altersdurchschnitt. Außerdem haben sich Männer in der Testgruppe eher für eine zusätzliche Technologie entschieden im Vergleich zu weiblichen TeilnehmerInnen.

Tabelle 11. Stichprobekennwerte der Variable Alter (in Jahren) aufgegliedert nach Untersuchungsgruppen in Studienphase 2.

Gruppe	*M (SD)*	*Mdn*	*Min*	*Max*
Kein Zusatz (*n* = 40)	72.27 (5.90)	72	61	89

Fortsetzung Tabelle 11.

Gruppe	*M (SD)*	*Mdn*	*Min*	*Max*
Telemedizinpaket (*n* = 31)	72.23 (7.38)	72	61	86
Sicherheitspaket (*n* = 11)	74.86 (7.29)	74	65	87
Vergleichsgruppe (*n* = 58)	72.82 (6.67)	73	60	88
Gesamt (*n* = 140)	72.64 (6.67)	72.50	60	89

Anmerkungen. M = Mittelwert. *SD* = Standardabweichung. *Mdn* = Median. *Min* = Minimum. *Max* = Maximum.

Tabelle 12. Relative Häufigkeiten der Variable Geschlecht der untersuchten Stichprobe in Studienphase 2.

	Relative Häufigkeit (in %)	
Gruppe	Frauen	Männer
Kein Zusatz (*n* = 40)	80.00	20.00
Telemedizinpaket (*n* = 31)	64.52	35.48
Sicherheitspaket (*n* = 11)	63.64	36.36
Vergleichsgruppe (*n* = 58)	75.86	24.14
Gesamt (*n* = 140)	73.58	26.43

Tabelle 13. Relative Häufigkeiten der Variable Bildungsstand (entspricht der höchsten abgeschlossenen Ausbildung) in der untersuchten Stichprobe in Studienphase 2.

	Relative Häufigkeit (in %)			
Gruppe	Pflicht-schulab-schluss	Lehr- / Fachschul-abschluss	Matura	Universitäts-abschluss
Kein Zusatz (*n* = 40)	0.00	40.00	40.00	20.00
Telemedizinpaket (*n* = 31)	9.68	38.71	32.26	19.35
Sicherheitspaket (*n* = 11)	9.00	27.27	36.36	27.27
Vergleichsgruppe (*n* = 58)	3.45	51.72	29.31	15.52
Gesamt (*n* = 140)	4.29	43.57	33.57	18.57

Tabelle 14. Relative Häufigkeiten der Variable Haushaltsgröße (entspricht der Anzahl von Personen in einem Haushalt) in der untersuchten Stichprobe in Studienphase 2.

	Relative Häufigkeit (in %)		
Gruppe	1 Person	2 Personen	>= 3 Personen
Kein Zusatz (*n* = 40)	40.00	57.50	2.50
Telemedizinpaket (*n* = 31)	61.29	32.26	6.48
Sicherheitspaket (*n* = 11)	54.54	45.46	0.00
Vergleichsgruppe (*n* = 58)	56.90	39.66	3.44
Gesamt (*n* = 140)	52.89	43.57	3.54

5 Ergebnisse

Zur Aufbereitung, Analyse und Visualisierung der Daten wurde R, Version 3.6.1, verwendet (R Core Team, 2019). Alle zusätzlich verwendeten R Packages sind im Literaturverzeichnis angeführt.

5.1 Studienphase 1

Dieser Abschnitt beschreibt zunächst Ergebnisse aus der ersten Studienphase bezüglich möglicher Einflüsse der Nutzung des WAALTeR-Systems auf ausgewählte Messgrößen („Wirkungsanalyse“). Es folgt eine Untersuchung des Nutzungsverhaltens und der Bewertung des WAALTeR-Systems (Tablet und Uhr).

5.1.1 Wirkungsanalyse

Die Wirkungsanalyse in Bezug auf die verschiedenen Messgrößen (Selbstbestimmtheit, Autonomie, Soziale Partizipation, Sicherheitsempfinden) während der ersten Studienphase umfasst mehrere Ansätze zur Modellierung: Erstens wird jeweils ein einfaches Modell geprüft, das ausschließlich Gruppenunterschiede in Bezug auf die jeweilige Messgröße zum letzten Messzeitpunkt (1.4) betrachtet. Unter der Voraussetzung, dass die TeilnehmerInnen zu Beginn der Untersuchung randomisiert zu Untersuchungsgruppen zugeteilt wurden, erlaubt diese Analyse eine globale Betrachtung möglicher Gruppenunterschiede zum abschließenden Messzeitpunkt der Studienphase 1. Zweitens wird je ein Modell gebildet, das die Veränderungen und Gruppenunterschiede über den Studienzeitraum beziehungsweise über die unterschiedlichen Messzeitpunkte berücksichtigt (in der Regel unter Verwendung von Linear Mixed Effects Model). Diese Modelle erlauben eine detaillierte Betrachtung von Gruppenunterschieden und die Veränderung von Messgrößen im Verlauf der Studie. Dabei wurden Untersuchungsgruppe und Messzeitpunkt (modelliert als Anzahl an Tagen seit dem ersten Messzeitpunkt) sowie deren Interaktion als Fixed Effects angenommen und Subjekt (individuelle StudienteilnehmerInnen) als Random Effect. In Bezug auf alle berichteten Linear Mixed Effects Models wurden Alter, Geschlecht und Haushaltgröße als Kovariaten berücksichtigt.[4] *p*-Werte in Bezug auf die jeweiligen Koeffizienten wurden unter Verwendung der Satterthwaite-Methode ermittelt[5] (Kuznetsova, Brockhoff, & Christensen, 2017; Giesbrecht & Burns, 1985; Fai & Cornelius, 1996). Alle *p*-Werte wurden adjustiert unter Verwendung der Korrekturmethode nach Benjamini und Yekutieli (2001). Diese Methode wurde aufgrund von Korrelationen zwischen abhängigen Variablen gewählt.

[4] Im Rahmen der Wirkungsanalyse wurde in einem initialen Prozess der Modellexploration die Verwendung verschiedener Kovariaten exploriert. Auf dieser Grundlage wurde die Auswahl von Kovariaten auf Alter, Geschlecht und Haushaltsgröße beschränkt, auch wenn weitere Kovariaten (z.B. physische Lebensqualität, Vorerfahrung mit Technologie) ebenfalls als angemessen betrachtet werden können.

[5] Die Berechnung von *p*-Werten in Bezug auf Linear Mixed Effects Models ist ein kontrovers diskutiertes Thema. Eine Vielzahl verschiedener Berechnungsmethoden existiert. Manche AutorInnen lehnen die Verwendung von *p*-Werten in diesem Kontext generell ab (siehe Johansson, 2011).

Muster in Bezug auf fehlende Werte wurden unter Verwendung der von Tierney et al. (2015) vorgeschlagene Methode analysiert, die Decision Trees verwendet, um die Struktur von Daten mit fehlenden Werten zu explorieren. Dabei zeigte sich, dass die Gruppenzugehörigkeit, der Messzeitpunkt und das Alter der Person starke Indikatoren für das Fehlen von Werten sind. Dieses Ergebnis ergibt intuitiv Sinn: Im Verlauf der Studie stiegen vor allem jüngere Personen der Vergleichsgruppe aus der Studie aus. Auf Basis inhaltlicher Analysen der Fragebogenitems kann angenommen werden, dass die Daten generell als Missing At Random (MAR) klassifiziert werden können. Der Argumentation von Little (1992) folgend wurden im Rahmen der Wirkungsanalyse fehlende Werte der jeweiligen abhängigen Variablen nicht imputiert.

Wie bereits erwähnt wurde - um die Unabhängigkeit der Datensätze zu gewährleisten – im Rahmen der Wirkungsanalyse in Bezug auf TeilnehmerInnen, die gemeinsam in einem Haushalt leben, jeweils nur eine Person pro Haushalt in der Auswertung berücksichtigt. Die jeweilige Person wurde vor der randomisierten Zuteilung zu den Untersuchungsgruppen zufällig ausgewählt.

5.1.1.1 Einflüsse auf die Selbstbestimmtheit

Das Konstrukt der Selbstbestimmtheit wurde in Anlehnung an die Studie von Dupuy et al. (2016) als eine zentrale Untersuchungsvariable gewählt. Wie in D5.1 beschrieben gliedert sich das psychologische Konstrukt der Selbstbestimmtheit – operationalisiert durch die BMPN (Neubauer & Voss, 2016) - in drei Subdimensionen:

- Autonomy
- Competence
- Relatedness

Zu bemerken ist, dass sich diese Subskalen inhaltlich mit Konstrukten überschneiden, die Gegenstand weiterer Fragestellungen sind. Insbesondere die BMPN-Dimension Autonomy ist inhaltlich sehr ähnlich zu dem Konstrukt Autonomie (im Rahmen der WAALTeR-Studie operationalisiert durch die WHOQOL-OLD Subskala Autonomie). Im Sinne der theoretischen Konsistenz werden diese sich inhaltlich überschneidenden Konstrukte immer in Bezug auf die jeweilige Skala betrachtet, durch die das Konstrukt operationalisiert ist. Tabelle 15 bis Tabelle 17 stellen deskriptive Stichprobenkennwerte zu den einzelnen Messzeitpunkten in Bezug auf die drei BMPN-Subskalen dar.

Tabelle 15. Stichprobekennwerte der BMPN-Subskala Autonomy aufgegliedert nach Untersuchungszeitpunkt und Untersuchungsgruppe.

	Testgruppe		Vergleichsgruppe	
Messzeitpunkt	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
1.0	5.77 (0.77)	5.67	5.52 (0.88)	5.50
1.1	5.68 (0.96)	5.83	5.60 (0.87)	5.67
1.2	5.62 (0.83)	5.67	5.33 (1.12)	5.33
1.3	5.60 (0.89)	5.67	5.59 (0.95)	5.67

Fortsetzung Tabelle 15.

	Testgruppe		Vergleichsgruppe	
Messzeitpunkt	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
1.4	5.67 (0.95)	5.67	5.34 (1.28)	5.67

Anmerkungen. M = Mittelwert; *SD* = Standardabweichung; *Mdn* = Median.

Tabelle 16. Stichprobekennwerte der BMPN-Subskala Competence aufgegliedert nach Untersuchungszeitpunkt und Untersuchungsgruppe.

	Testgruppe		Vergleichsgruppe	
Messzeitpunkt	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
1.0	5.41 (0.98)	5.33	5.48 (0.89)	5.50
1.1	5.42 (0.93)	5.50	5.44 (1.01)	5.67
1.2	5.37 (0.92)	5.33	5.33 (0.99)	5.50
1.3	5.24 (1.04)	5.17	5.18 (0.97)	5.25
1.4	5.34 (0.98)	5.17	5.34 (1.05)	5.50

Anmerkungen. M = Mittelwert; *SD* = Standardabweichung; *Mdn* = Median.

Tabelle 17. Stichprobekennwerte der BMPN-Subskala Relatedness aufgegliedert nach Untersuchungszeitpunkt und Untersuchungsgruppe.

	Testgruppe		Vergleichsgruppe	
Messzeitpunkt	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
1.0	5.71 (0.92)	5.83	5.73 (0.83)	5.83
1.1	5.76 (0.83)	5.83	5.68 (0.93)	5.83
1.2	5.65 (0.93)	5.83	5.46 (1.04)	5.67
1.3	5.63 (0.95)	5.67	5.68 (1.01)	6.00
1.4	5.66 (0.97)	5.67	5.62 (0.94)	5.67

Anmerkungen. M = Mittelwert; *SD* = Standardabweichung; *Mdn* = Median.

Zur Überprüfung der Gruppenunterschiede in Bezug auf die drei BMPN-Subskalen zum abschließenden Messzeitpunkt der ersten Studienphase (1.4) wurden *t*-Tests für unabhängige Stichproben durchgeführt. Die Unterschiede zwischen den Untersuchungsgruppen in Bezug auf die drei Subskalen zum Messzeitpunkt 1.4 sind nicht signifikant. Tabelle 18 beschreibt die Kennwerte der jeweiligen *t*-Tests.

Tabelle 18. *t*-Test-Kennwerte zum Vergleich der beiden Untersuchungsgruppen hinsichtlich der BMPN-Subskalen am abschließenden Messzeitpunkt (1.4).

BMPN-Subskala	*t*	*df*	*p*	*d*
Autonomy	1.45	67.63	ns	0.30

Fortsetzung Tabelle 18.

BMPN-Subskala	*t*	*df*	*p*	*d*
Competence	-0.01	80.84	ns	0.00
Relatedness	0.25	88.02	ns	0.05

Anmerkungen. n = 114. *t* = Prüfgröße. *df* = Freiheitgrade. *p* = Signifikanzwert. *d* = Effektstärke nach Cohen (1988). ns = nicht signifikant.

Wie in Tabelle 18 zu erkennen unterscheiden sich die Untersuchungsgruppen hinsichtlich der drei BMPN-Subskalen zum abschließenden Messzeitpunkt (1.4) nicht signifikant voneinander. Bemerkenswert ist allerdings, dass die Daten in Bezug auf Subskala Autonomy auf einen kleinen (nicht signifikanten) Effekt (d = 0.30) hinweisen: Personen aus der Testgruppe berichteten höhere Werte zum abschließenden Messzeitpunkt.

Um mögliche Veränderungen der einzelnen BMPN-Subskalen über die Zeit und zwischen Gruppen zu untersuchen, wurde eine Analyse mittels Linear Mixed Effects Models durchgeführt. Tabelle 19 bis Tabelle 21 stellen die Parameterschätzungen bezüglich der untersuchungsrelevanten[6] Fixed Effects des jeweiligen Modells für die drei BMPN-Subskalen dar. Es zeigen sich auf Grundlage der Linear Mixed Effects Models keinerlei signifikanten Effekte von Untersuchungsgruppe und Messzeitpunkt (beziehungsweise deren Interaktion).

Tabelle 19. Parameterschätzungen für die untersuchungsrelevanten Fixed Effects des Linear Mixed Effects Models für die BMPN-Subskala Autonomy.

Effekt	Est	*SE*	Est (st)	*SE* (st)	*df*	*t*	*p*
UG (VG)	-0.20	0.15	-0.10	0.07	139.93	-1.29	ns
MZP	-0.00	0.00	-0.05	0.03	201.43	-1.42	ns
MZP * UG	0.00	0.00	0.02	0.05	478.07	0.48	ns

Anmerkungen. Alle Kennwerte gerundet auf zwei Nachkommastellen. *n* = 141. Konstante = 5.92. Est = Parameterschätzung. *SE* = Standardfehler. st = standardisiert. *df* = Freiheitgrade. *t* = Prüfgröße. *p* = Signifikanzwert. ns = nicht signifikant. UG = Untersuchungsgruppe. VG = Vergleichsgruppe. MZP = Messzeitpunkt. Kovariaten im Modell: Geschlecht (nicht signifikant), Alter (nicht signifikant), Haushaltgröße (nicht signifikant). Messzeitpunkt modelliert als Anzahl an Tagen seit dem ersten Messzeitpunkt (1.0).

Tabelle 20. Parameterschätzungen für die untersuchungsrelevanten Fixed Effects des Linear Mixed Effects Models für die BMPN-Subskala Competence.

Effekt	*Est*	*SE*	*Est (st)*	*SE (st)*	*df*	*t*	*p*
UG (VG)	0.07	0.16	0.04	0.08	191.69	0.48	ns

[6] „Untersuchungsrelevant“ bezeichnet in diesem Zusammenhang die Variablen Untersuchungsgruppe, Messzeitpunkt sowie deren Interaktion. Kovariaten sind in diesem Sinn keine untersuchungsrelevanten Variablen.

Fortsetzung Tabelle 20.

Effekt	Est	*SE*	Est (st)	*SE* (st)	*df*	*t*	*p*
MZP	0.00	0.00	-0.04	0.03	470.46	-1.28	ns
MZP * UG	-0.00	0.00	-0.05	0.04	479.60	-1.25	ns

Anmerkungen. Alle Kennwerte gerundet auf zwei Nachkommastellen. *n* = 141. Konstante = 6.88. Est = Parameterschätzung. *SE* = Standardfehler. st = standardisiert. *df* = Freiheitgrade. *t* = Prüfgröße. *p* = Signifikanzwert. ns = nicht signifikant. UG = Untersuchungsgruppe. VG = Vergleichsgruppe. MZP = Messzeitpunkt. Kovariaten im Modell: Geschlecht (nicht signifikant), Alter (nicht signifikant), Haushaltgröße (nicht signifikant). Messzeitpunkt modelliert als Anzahl an Tagen seit dem ersten Messzeitpunkt (1.0).

Tabelle 21. Parameterschätzungen für die untersuchungsrelevanten Fixed Effects des Linear Mixed Effects Models für die BMPN-Subskala Relatedness.

Effekt	Est	*SE*	Est (st)	*SE* (st)	*df*	*t*	*p*
UG (VG)	-0.04	0.15	-0.02	0.08	212.48	-0.30	ns
MZP	-0.00	0.00	-0.04	0.03	472.42	-1.03	ns
MZP * UG	0.00	0.00	0.00	0.00	481.73	0.06	ns

Anmerkungen. Alle Kennwerte gerundet auf zwei Nachkommastellen. *n* = 141. Konstante = 5.39. Est = Parameterschätzung. *SE* = Standardfehler. st = standardisiert. *df* = Freiheitgrade. *t* = Prüfgröße. *p* = Signifikanzwert. ns = nicht signifikant. UG = Untersuchungsgruppe. VG = Vergleichsgruppe. MZP = Messzeitpunkt. Kovariaten im Modell: Geschlecht (nicht signifikant), Alter (nicht signifikant), Haushaltsgröße (nicht signifikant). Messzeitpunkt modelliert als Anzahl an Tagen seit dem ersten Messzeitpunkt (1.0).

5.1.1.2 Einflüsse auf die Autonomie

Wie in D5.1 beschrieben wurde die wahrgenommene Autonomie der TeilnehmerInnen mit der Subskala Autonomie des World Health Organisation Quality of Life OLD operationalisiert (WHOQOL-OLD; Conrad et al. 2016). Tabelle 22 stellt Stichprobekennwerte der Subskala aufgegliedert nach Untersuchungszeitpunkt und -gruppe dar.

Tabelle 22. Stichprobekennwerte der WHOQOL-OLD-Subskala Autonomie aufgegliedert nach Untersuchungszeitpunkt und Untersuchungsgruppe.

	Testgruppe		Vergleichsgruppe	
Messzeitpunkt	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
1.0	75.24 (12.96)	75.00	74.41 (13.71)	75.00
1.1	73.52 (13.57)	75.00	73.85 (14.75)	75.00
1.2	74.75 (14.67)	75.00	74.42 (16.53)	75.00
1.3	75.35 (14.46)	75.00	74.84 (15.40)	75.00
1.4	75.79 (15.74)	75.00	73.96 (16.67)	75.00

Anmerkungen. M = Mittelwert. *SD* = Standardabweichung. *Mdn* = Median.

Zur Untersuchung der Gruppenunterschiede zum abschließenden Messzeitpunkt wurde ein *t*-Test für unabhängige Stichproben durchgeführt. Auch in Bezug auf die Autonomie zeigen sich keine signifikanten Unterschiede zwischen den Untersuchungsgruppen zum abschließenden Messzeitpunkt (siehe Tabelle 23).

Tabelle 23. *t*-Test-Kennwerte zum Vergleich der beiden Untersuchungsgruppen hinsichtlich der WHOQOL-OLD-Subskala Autonomie am abschließenden Messzeitpunkt (1.4).

	t	*df*	*p*	*d*
Autonomie	*0.58*	*82.22*	ns	*0.11*

Anmerkungen. n = 113. *t* = Testprüfgröße des *t*-Tests. *df* = Freiheitgrade. *p* = Signifikanzwert. *d* = Effektstärke nach Cohen (1988). ns = nicht signifikant.

Um mögliche Veränderungen der Autonomie über die Zeit und zwischen Gruppen zu untersuchen, wurde wiederum eine Analyse mittels Linear Mixed Effects Models durchgeführt. Tabelle 24 stellt Parameterschätzungen bezüglich der untersuchungsrelevanten Fixed Effects des Modells für Autonomie dar. Es zeigen sich keine signifikanten Effekte von Untersuchungsgruppe und Messzeitpunkt (beziehungsweise deren Interaktion).

Tabelle 24. Parameterschätzungen für die untersuchungsrelevanten Fixed Effects des Linear Mixed Effects Models für WHOQOL-OLD-Subskala Autonomie.

Effekt	Est	*SE*	Est (st)	*SE* (st)	*df*	*t*	*p*
UG (VG)	-0.24	2.34	-0.01	0.08	182.98	-0.10	ns
MZP	*0.01*	*0.00*	*0.05*	*0.03*	*460.51*	*1.66*	ns
MZP * UG	-0.00	0.00	-0.03	0.04	467.45	-0.84	ns

Anmerkungen. Alle Kennwerte gerundet auf zwei Nachkommastellen. *n* = 141. Konstante = 97.46. Est = Parameterschätzung. *SE* = Standardfehler. st = standardisiert. *df* = Freiheitgrade. *t* = Prüfgröße. *p* = Signifikanzwert. ns = nicht signifikant. UG = Untersuchungsgruppe. VG = Vergleichsgruppe. MZP = Messzeitpunkt. Kovariaten im Modell: Geschlecht (nicht signifikant), Alter (nicht signifikant), Haushaltgröße (nicht signifikant). Messzeitpunkt modelliert als Anzahl an Tagen seit dem ersten Messzeitpunkt (1.0).

5.1.1.3 Einflüsse auf die soziale Partizipation

Wie auch bereits in D5.1 beschrieben wurde die soziale Partizipation der TeilnehmerInnen durch die Facette Soziale Partizipation des WHOQOL-OLD operationalisiert (Conrad et al. 2016). Tabelle 25 stellt Stichprobekennwerte der Subskala aufgegliedert nach Untersuchungszeitpunkt und -gruppe dar.

Tabelle 25. Stichprobekennwerte der WHOQOL-OLD-Subskala Soziale Partizipation aufgegliedert nach Untersuchungszeitpunkt und Untersuchungsgruppe.

	Testgruppe		Vergleichsgruppe	
Messzeitpunkt	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
1.0	74.50 (15.97)	75.00	74.13 (16.17)	75.00
1.1	75.48 (15.79)	75.00	77.08 (13.97)	78.13

Fortsetzung Tabelle 25.

	Testgruppe		Vergleichsgruppe	
Messzeitpunkt	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
1.2	74.40 (16.56)	75.00	75.00 (13.84)	75.00
1.3	75.18 (17.32)	78.13	74.68 (15.10)	75.00
1.4	76.01 (16.68)	78.13	76.52 (15.80)	75.00

Anmerkungen. M = Mittelwert. *SD* = Standardabweichung. *Mdn* = Median.

Zur Untersuchung der Gruppenunterschiede zum abschließenden Messzeitpunkt wurde ein *t*-Test für unabhängige Stichproben durchgeführt. Die Unterschiede zwischen den Gruppen hinsichtlich Sozialer Partizipation sind nicht signifikant (siehe Tabelle 26).

Tabelle 26. *t*-Test-Kennwerte zum Vergleich der beiden Untersuchungsgruppen hinsichtlich der WHOQOL-Subskala Soziale Partizipation am abschließenden Messzeitpunkt (1.4).

	t	*df*	*p*	*d*
Soziale Partizipation	*-0.16*	*88.07*	ns	*0.03*

Anmerkungen. n = 113. *t* = Prüfgröße des *t*-Tests. *df* = Freiheitgrade. *p* = Signifikanzwert. *d* = Effektstärke nach Cohen (1988). ns = nicht signifikant.

Des Weiteren wurde eine Analyse mittels Linear Mixed Effects Models durchgeführt. Zusammenfassend zeigten sich keine signifikanten Effekte von Untersuchungsgruppe, Messzeitpunkt und deren Interaktion auf Soziale Partizipation (sieheTabelle 27).

Tabelle 27. Parameterschätzungen für die untersuchungsrelevanten Fixed Effects des Linear Mixed Effects Models für die WHOQOL-OLD-Subskala Soziale Partizipation.

Effekt	Est	*SE*	Est (st)	*SE* (st)	*df*	*t*	*p*
UG (VG)	1.92	2.61	0.06	0.08	178.23	0.73	ns
MZP	0.01	0.00	0.05	0.03	443.97	1.55	ns
MZP * UG	-0.01	0.01	-0.04	0.04	450.12	-1.02	ns

Anmerkungen. Alle Kennwerte gerundet auf zwei Nachkommastellen. *n* = 141. Konstante = 98.79. Est = Parameterschätzung. *SE* = Standardfehler. st = standardisiert. *df* = Freiheitgrade. *t* = Prüfgröße. *p* = Signifikanzwert. ns = nicht signifikant. UG = Untersuchungsgruppe. VG = Vergleichsgruppe. MZP = Messzeitpunkt. Kovariaten im Modell: Geschlecht (nicht signifikant), Alter (nicht signifikant), Haushaltgröße (nicht signifikant). Messzeitpunkt modelliert als Anzahl an Tagen seit dem ersten Messzeitpunkt (1.0).

5.1.1.4 Einflüsse auf das Sicherheitsempfinden

Das Sicherheitsempfinden wurde mit einer selbstentwickelten Skala mit fünf Items erhoben. Hohe Ausprägungen auf der Skala entsprechen einem hohen Sicherheitsgefühl. Tabelle 28 stellt Stichprobenkennwerte in Bezug auf das Sicherheitsempfinden nach Untersuchungszeitpunkt und -gruppe dar.

Tabelle 28. Stichprobekennwerte in Bezug auf das Sicherheitsempfinden aufgegliedert nach Untersuchungszeitpunkt und Untersuchungsgruppe.

	Testgruppe		Vergleichsgruppe	
Messzeitpunkt	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
1.0	5.26 (0.61)	5.25	5.19 (0.55)	5.25
1.1	5.20 (0.55)	5.25	5.09 (0.56)	5.00
1.2	5.21 (0.56)	5.25	5.03 (0.63)	5.00
1.3	5.01 (0.66)	5.25	5.08 (0.65)	5.25
1.4	5.19 (0.60)	5.25	5.02 (0.60)	5.00

Anmerkungen. M = Mittelwert. *SD* = Standardabweichung. *Mdn* = Median.

Des Weiteren stellt Tabelle 29 Kennwerte des *t*-Tests in Bezug auf das Sicherheitsempfinden zum abschließenden Messzeitpunkt (1.4) dar. Obwohl die Effektstärke (*d* = 0.27) auf einen kleinen Effekt hinweist, sind die Unterschiede zwischen Gruppen nicht signifikant.

Tabelle 29. *t*-Test-Kennwerte zum Vergleich der beiden Untersuchungsgruppen hinsichtlich des Sicherheitsempfindens am abschließenden Messzeitpunkt (1.4).

	t	*df*	*p*	*d*
Sicherheitsempfinden	-1.41	88.80	ns	0.27

Anmerkungen. *n* = 113. *t* = Testprüfgröße des *t*-Tests. *df* = Freiheitgrade. *p* = Signifikanzwert. *d* = Effektstärke nach Cohen (1988). ns = nicht signifikant.

Des Weiteren wurde eine Analyse mittels Linear Mixed Effects Models durchgeführt. Tabelle 30 stellt Parameterschätzungen für untersuchungsrelevante Fixed Effects bezüglich des entsprechenden Modells dar. Zusammenfassend zeigten sich keine signifikanten Effekte von Untersuchungsgruppe, Messzeitpunkt und deren Interaktion auf das Sicherheitsempfinden.

Tabelle 30. Parameterschätzungen für die untersuchungsrelevanten Fixed Effects des Linear Mixed Effects Models für Sicherheitsempfinden.

Effekt	Est	*SE*	Est (st)	*SE* (st)	*df*	*t*	*p*
UG (VG)	-0.08	0.09	-0.07	0.07	254.18	-0.92	ns
MZP	-0.00	0.00	-0.05	0.03	486.57	-1.25	ns
MZP * UG	-0.00	0.00	-0.01	0.06	496.39	-0.18	ns

Anmerkungen. Alle Kennwerte gerundet auf zwei Nachkommastellen. *n* = 141. Konstante = 5.39. Est = Parameterschätzung. *SE* = Standardfehler. st = standardisiert. *df* = Freiheitgrade. *t* = Prüfgröße. *p* = Signifikanzwert. ns = nicht signifikant. UG = Untersuchungsgruppe. VG = Vergleichsgruppe. MZP = Messzeitpunkt. Kovariaten im Modell: Geschlecht (nicht signifikant), Alter (nicht signifikant), Haushaltgröße (nicht signifikant). Messzeitpunkt modelliert als Anzahl an Tagen seit dem ersten Messzeitpunkt (1.0).

5.1.1.5 Weitere Anmerkungen zur Wirkungsanalyse

Obwohl keiner der untersuchten Effekte Signifikanz erreichte, weisen die Daten teilweise auf kleine Effekte hin (z.B. *d >= 0.2).* Die Richtung der Effekte ist dabei erwartungskonform (z.B. höheres Sicherheitsempfinden in der Testgruppe zum abschließenden Messzeitpunkt). Auf Basis von Poweranalysen erscheint die Power des Untersuchungsdesigns beziehungsweise der verwendeten statistischen Verfahren nicht ausreichend, um solch kleinen Effekte zu detektieren. Damit soll nicht impliziert werden, dass im Fall einer größeren Stichprobe ein gegebener Effekt jedenfalls signifikant gewesen wäre. Stattdessen ist diese Beobachtung ein Hinweis darauf, dass Effekte durch Technologien wie WAALTeR relativ klein sein könnten, was bei der Planung zukünftiger Studien Berücksichtigung finden kann.

Zusätzlich zu den vorher beschriebenen Auswertungen wurden auf Basis explorativer Analysen auch bestimmte Subgruppen der Stichprobe untersucht (Personen über dem Medianalter, Personen ohne Vorerfahrung mit Smartphone beziehungsweise Tablet). Die Ergebnisse aus diesen Analysen weisen darauf hin, dass bei Personen mit geringer Vorerfahrung mit Technologie zum abschließenden Messzeitpunkt tendenziell größere Gruppenunterschiede hinsichtlich dem Sicherheitsempfinden und der BMPN-Skala Autonomy existieren als in Bezug auf die Gesamtstichprobe. Diese Unterschiede sind allerdings ebenfalls nicht signifikant und sollten im Allgemeinen mit Vorsicht interpretiert. Sie können allerdings ein Orientierungspunkt für zukünftige Forschung sein.

5.1.2 Nutzung des WAALTeR-Systems

In diesem Abschnitt wird die Nutzung des WAALTeR-Systems exploriert. Für diese Analyse wurde eine vorwiegend deskriptive Vorgehensweise gewählt. Der Fokus liegt hierbei auf Veränderungen der Nutzung über die Zeit, Unterschiede in der Nutzung verschiedener Anwendungen sowie Unterschiede in der individuellen Nutzung durch die StudienteilnehmerInnen. Die Nutzung des Systems kann Rückschlüsse auf die Präferenzen und Bedürfnisse der NutzerInnen erlauben und dabei helfen Verbesserungspotentiale zu identifizieren.

5.1.2.1 Nutzung des WAALTeR-Tablets

Wie bereits erwähnt wurde die Nutzung des WAALTeR-Tablets automatisiert mithilfe einer Tracking-Software erhoben. Aufgrund von Verzögerungen bei der technischen Implementierung konnten die Nutzungsdaten erst ab 08. Juni 2018 aufgezeichnet werden (ungefähr zwei Monate nach Start der Feldstudie). Dabei wurde auf aufgezeichnet, welche Anwendungen wie häufig von den einzelnen NutzerInnen aufgerufen wurden. WAALTeR-Anwendungen bezeichnen jene Anwendungen, die Rahmen von WAALTeR von Projektpartnern implementiert wurden. Zusätzliche Anwendungen bezeichnen hingegen Applikationen, die von Anbietern außerhalb des Projekts stammen und zusätzlich zu den WAALTeR-Anwendungen auf dem Tablet installiert wurden, um die Funktionalität des Tablets in sinnvoller Weise zu erweitern.[7]

[7] Für eine ausführliche Beschreibung des nutzerInnenzentrierten Designprozesses im Rahmen von WAALTeR siehe D2.1 / 2.2.

WAALTeR-Anwendungen am Tablet umfassen folgende Applikationen:

- (Einstellungen), FragNebenan, Gesundheitsdienste, Handbuch, Kalender, Messaging, Photocloud, Veranstaltungen, Wegplaner, Zeitungen

Zusätzliche Anwendungen am Tablet umfassen folgende Applikationen:

- Amazon, Apo-App, BILLA Onlineshop, E-Mail, Internetbrowser, Kindle, Mahjong, ÖBB App, ORF TVthek, Qando, Skype, Solitaire, Sudoku, WetterOnline, WhatsApp, Willhaben

Neben der Nutzungshäufigkeit (Anzahl der Aufrufe) wurde auch die Nutzungsdauer aufgezeichnet. Allerdings war diese Form der Aufzeichnung technisch nur in Bezug auf die WAALTeR-Anwendungen (nicht die zusätzlichen Anwendungen) möglich. Dies ist ein Grund, weshalb das Hauptaugenmerk der Analyse des Nutzungsverhaltens auf der Nutzungshäufigkeit liegt. Außerdem lässt sich annehmen, dass die Nutzungshäufigkeit eine bessere Vergleichbarkeit zwischen verschiedenen Anwendungen ermöglicht: Beispielsweise verbringen NutzerInnen vermutlich mehr Zeit mit einer Zeitungs-App als mit einer Wegplaner-App, obwohl die Nutzungshäufigkeit der beiden Apps womöglich ähnlich hoch ist. In diesem Sinne ist die Nutzungshäufigkeit wohl ein besserer Indikator für die „Beliebtheit“ einer bestimmten Anwendung.

Abbildung 1 stellt die Gesamtanzahl von Aufrufen pro Monat im Verlauf der ersten Studienphase dar. Dabei ist ein Abfall der Nutzungshäufigkeit im Studienverlauf erkennbar. Dabei fallen Monate, in denen Stammtische stattgefunden haben (Juli, Oktober, Januar), durch eine durch eine vergleichsweise hohe Nutzung auf.

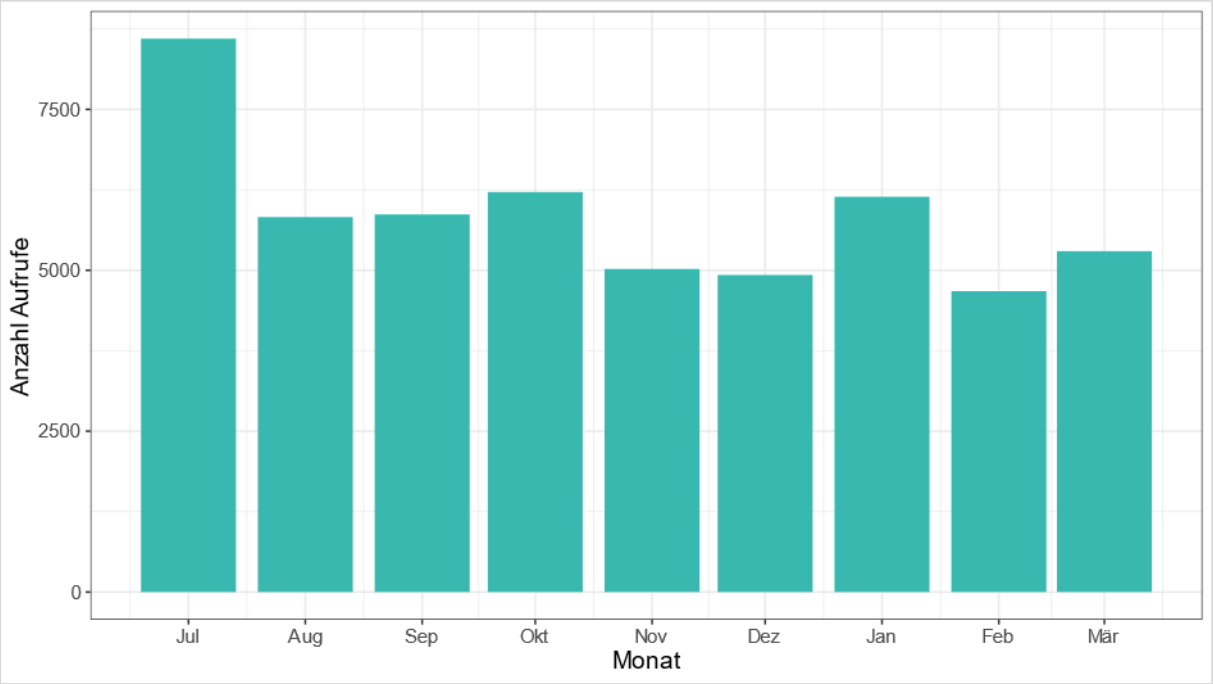

Abbildung 1. Gesamtanzahl von Anwendungsaufrufen pro Monat in Studienphase 1. Vormonate der Studienphase 1 aufgrund unvollständiger Daten nicht dargestellt.

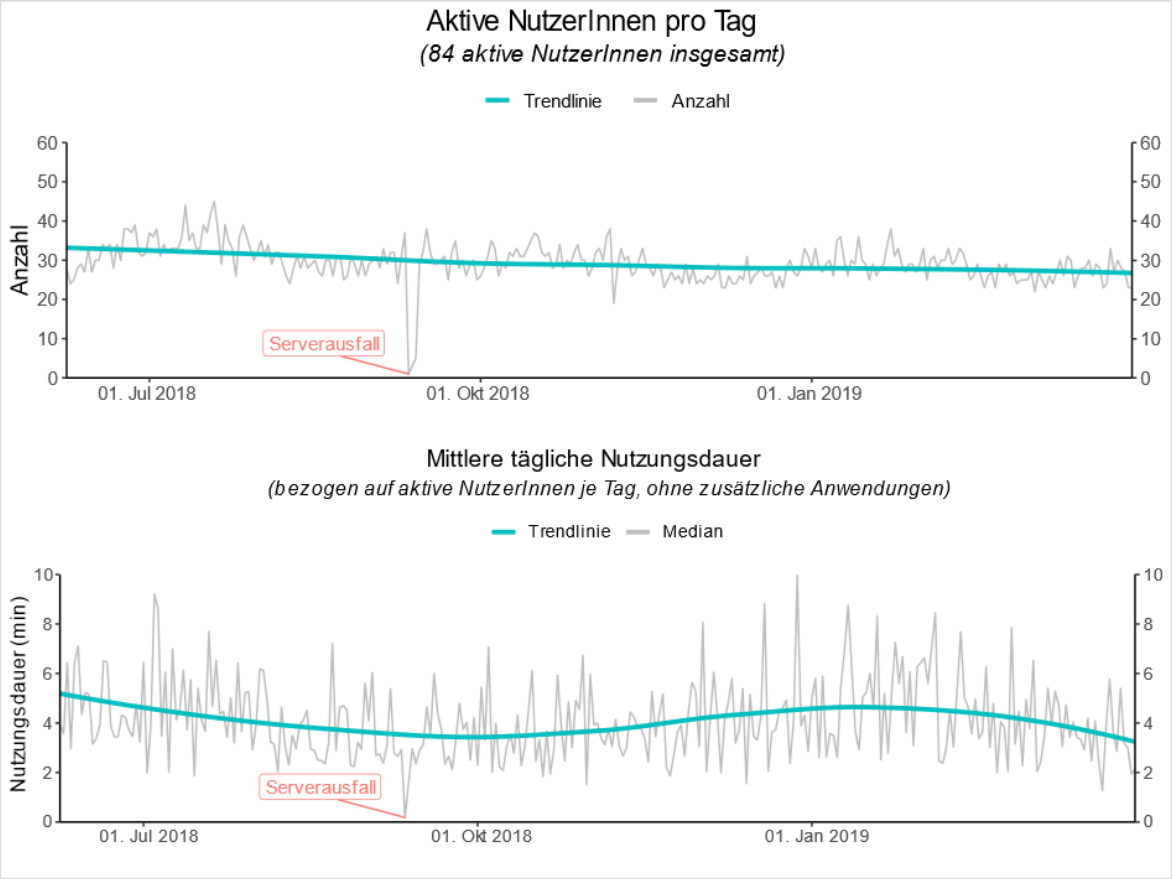

Abbildung 2. Aktive NutzerInnen pro Tag und mittlere tägliche Nutzungsdauer über die Zeit (Studienphase 1).

Die Tendenz einer sinkenden Nutzung im Studienverlauf zeigt sich auch in der Nutzungsdauer und der Anzahl individuellen NutzerInnen pro Tag (siehe Abbildung 2). Obwohl die mittlere tägliche Nutzungsdauer eine hohe Varianz aufweist, lässt die Trendlinie insgesamt ein Abfallen erkennen. In Bezug auf die Anzahl individueller NutzerInnen ist dieser Trend schwächer ausgeprägt. Die Gesamtanzahl individueller NutzerInnen von 84 Personen zeigt außerdem, dass einige TeilnehmerInnen der Testgruppe das Tablet überhaupt nicht genutzt haben (im Zeitraum der Aufzeichnung). Diese NutzerInnen gaben verschiedene Gründe für Nichtnutzung an: Häufig besäßen die Personen schon eine vergleichbare Technologie (z.B. Smartphone oder Laptop) und würde diese gegenüber dem WAALTeR-Tablet bevorzugen.

Zum Vergleich der Nutzung zwischen unterschiedlichen Anwendungen stellt Abbildung 3 Nutzungshäufigkeiten für die verschiedenen Anwendungen am Tablet dar (die zwölf Anwendungen mit den meisten Aufrufen insgesamt). Dabei fällt auf, dass zusätzliche Anwendungen im Vergleich zu den WAALTeR-Anwendungen insgesamt häufiger genutzt werden. Wenig überraschend ist wohl, dass Anwendungen wie der Webbrowser („Internet“) und E-Mail relativ oft genutzt werden. Darüber hinaus werden Spielen (Solitaire, Mahjong, Sudoku) vergleichsweise häufig genutzt. In Bezug auf die WAALTeR-Anwendungen hat die Zeitungsapp mit Abstand am meisten Zugriffe.

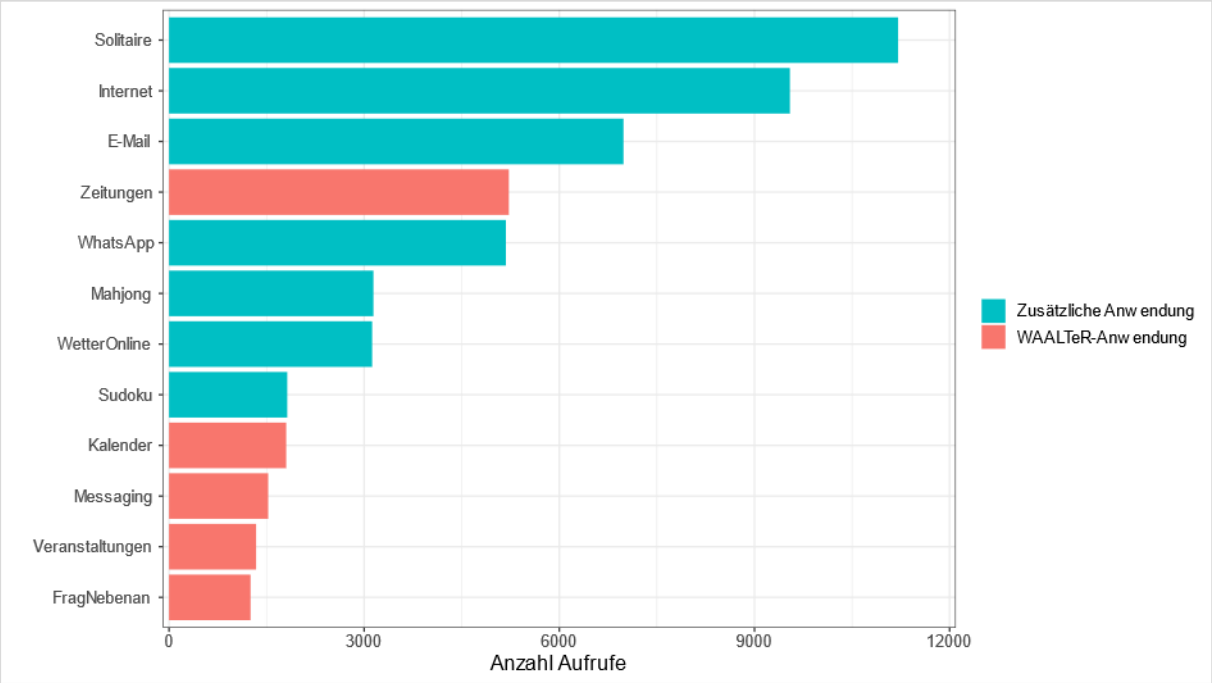

Abbildung 3. Die zwölf Anwendungen mit den meisten Aufrufen (Studienphase 1; Daten von 08. Juni 2018 bis 31. März 2019).

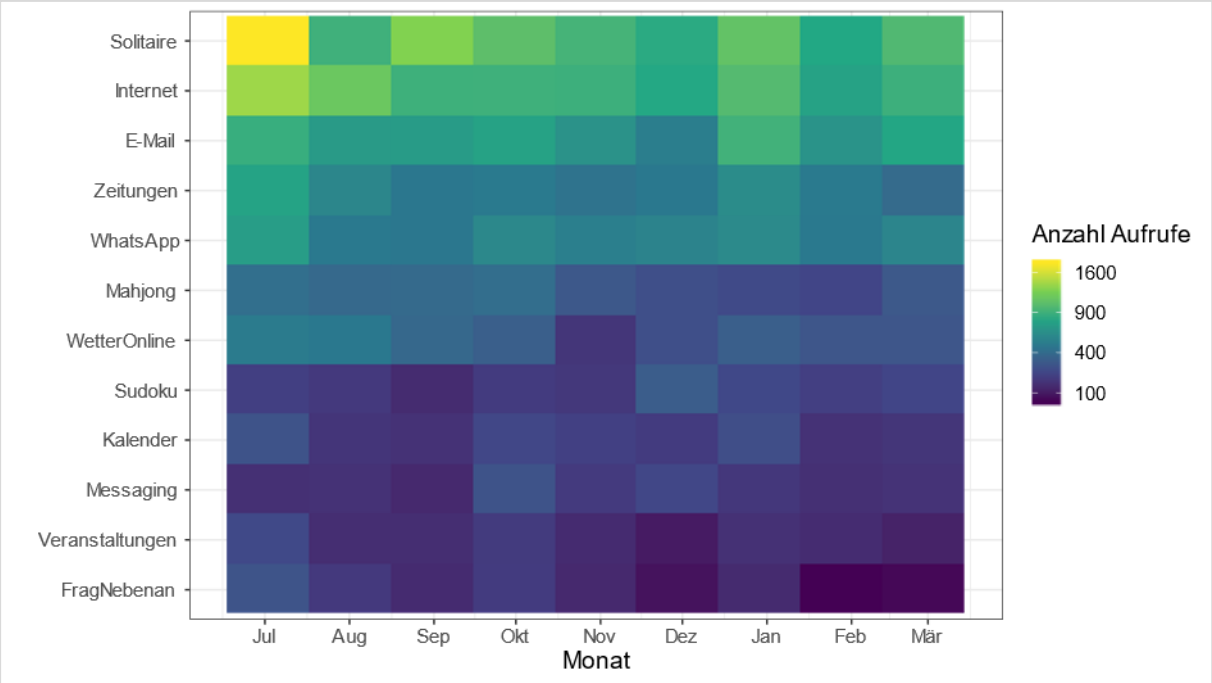

Abbildung 4. Anzahl der Aufrufe der 12 meistgenutzten Anwendungen aufgegliedert nach Monaten. Vormonate der Studienphase 1 aufgrund unvollständiger Daten nicht dargestellt.

Abbildung 4 führt die Darstellungen aus den Abbildungen 1 und 3 zusammen. Gut erkennbar ist auch hier ein zeitlicher Trend: Die allgemeine Nutzungshäufigkeit ist im Studienverlauf ins-

gesamt gesunken. Hinsichtlich der einzelnen Anwendungen gibt es weitere interessante Muster: Beispielsweise wurde Sudoku in den späteren Monaten der Studienphase häufiger genutzt als zu Beginn.

Allgemein ist zu bedenken, dass eine hohe Nutzungshäufigkeit nicht notwendigerweise mit einer hohen Anzahl individueller NutzerInnen einhergeht. So ist es beispielsweise auch denkbar, dass einige wenige NutzerInnen eine App sehr viel nutzen, wohingegen andere Personen die App kaum beziehungsweise gar nicht verwenden. Diese Annahme wird Zusammenhang mit der individuellen Nutzung näher betrachtet. Abbildung 5 stellt die Verteilung der relativen Nutzungshäufigkeit nach individuellen NutzerInnen dar. Relative Nutzungshäufigkeit entspricht dabei Anzahl an Tagen, an denen das Tablet von der jeweiligen Person genutzt wurde, im Verhältnis zur Gesamtanzahl an Studientagen. Etwa 50% aller TeilnehmerInnen nutzte das Tablet an weniger als 25% der Studientage (entspricht < 25% relativer Nutzungshäufigkeit). Die 25% der TeilnehmerInnen mit der höchsten Nutzungshäufigkeit nutzen das Tablet an mindestens 60% der Studientage (entspricht >= 60% relativer Nutzungshäufigkeit). Nur eine Handvoll NutzerInnen (ca. 10%) verwendete das Tablet an 90% der Studientage oder öfter. Dieses Ergebnis impliziert, dass das WAALTeR-Tablet (im „Durchschnitt“) keine Technologie ist, die von NutzerInnen tagtäglich Verwendung findet, sondern eher gelegentlich in die Hand genommen wird.

Des Weiteren stellt Abbildung 6 die Anzahl der Aufrufe der zwölf meistgenutzten Anwendungen aufgegliedert nach individuellen NutzerInnen dar. Punkte in der Abbildung entsprechen individuellen NutzerInnen (pro App). Die Größe der Punkte entspricht dabei der Anzahl an individuellen NutzerInnen pro App. Beispielweise zeichnet sich die Zeitungs-Anwendung durch eine hohe Anzahl individueller NutzerInnen aus (> 80 NutzerInnen), während Messaging von vergleichsweise wenigen Personen verwendet wurde. Gut erkennbar ist zum Beispiel, dass Solitaire (die Anwendung mit den meisten Aufrufen insgesamt) von relativ wenigen Personen verwendet wurden. Stattdessen ergibt sich die hohe Anzahl an Aufrufen durch eine wenige „Power-UserInnen“. Diese Tendenz (BenutzerInnen mit sehr vielen Aufrufen bei einer geringen Anzahl individueller NutzerInnen) ist in Bezug auf die Messaging-App besonders ausgeprägt. Ebenso gibt es Anwendungen wie die Veranstaltungs-App, die von vielen individuellen Nutzern verwendet wurde. Jedoch ist die Anzahl an Aufrufen pro NutzerIn relativ gering, was sich in einer niedrigeren Gesamtnutzungshäufigkeit widerspiegelt. Schließlich gibt es Anwendungen mit vielen individuellen NutzerInnen und auch einer hohen Gesamtanzahl von Aufrufen. Dies ist – wenig überraschend – beim Webbrowser („Internet“) der Fall.

Insgesamt zeigt die deskriptive Analyse der individuellen Nutzungshäufigkeit, dass starke Unterschiede im Nutzungsverhalten zwischen Personen existieren. Obwohl es einige Anwendungen gibt, die von vielen Personen genutzt wurden, schwankt die individuelle Nutzungshäufigkeit in Bezug auf die unterschiedlichen Anwendungen erheblich. Diese Heterogenität der Nutzung könnte einer der Gründe dafür sein, dass im Rahmen der Wirkungsanalyse keine signifikanten Unterschiede zwischen Gruppen und Messzeitpunkten gefunden wurden (siehe auch Diskussion). Einschränkend ist zu bemerken, dass die Aufzeichnung der Nutzung aufgrund

von Verzögerungen bei der technischen Implementierung erst ca. zwei Monate nach Studienstart begann, sodass das Nutzungsverhalten in diesem Anfangszeitraum nicht weiter untersucht werden konnte (Gesamtdauer der ersten Studienphase: zwölf Monate).

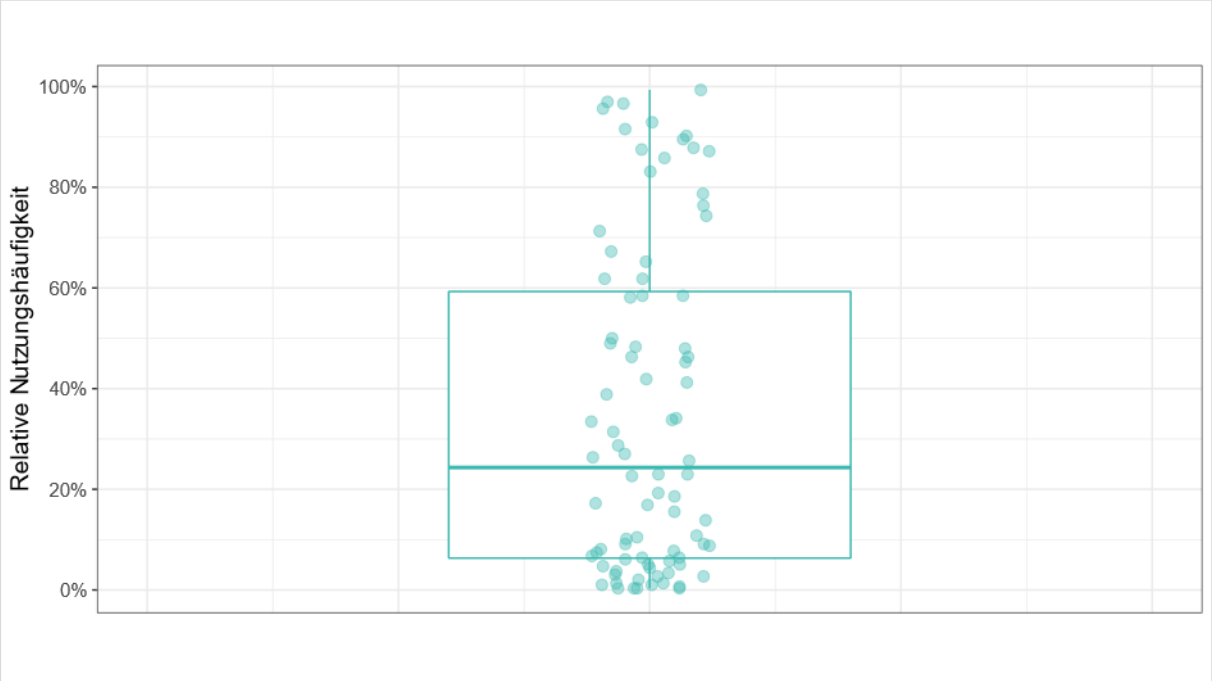

Abbildung 5. Verteilung der relativen Nutzungshäufigkeit pro Person. Relative Nutzungshäufigkeit entspricht den Tagen, an denen das Tablet genutzt wurde, im Verhältnis zur Gesamtanzahl an Studientagen (Tage mit fehlenden Daten ausgeschlossen).

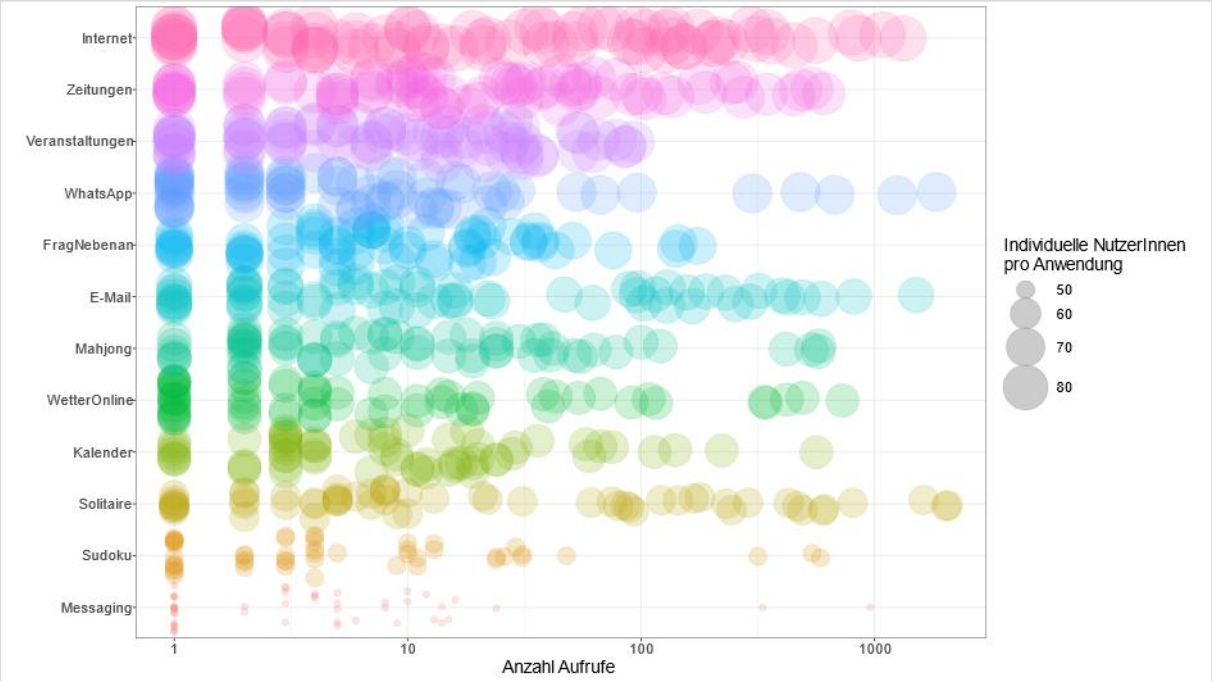

Abbildung 6. Anzahl der Aufrufe der zwölf meistgenutzten Anwendungen aufgegliedert nach individuellen NutzerInnen. Punkte in der Abbildung entsprechen individuellen NutzerInnen. Die Anzahl der Aufrufe ist auf einer logarithmischen Skala dargestellt.

5.1.2.2 Nutzung der WAALTeR-Uhr

Die Nutzung der WAALTeR-Uhr wurde mittels Fragebogen erhoben. Dazu beantworteten TeilnehmerInnen zu jedem Messzeitpunkt folgende Fragen:

- *Wie häufig tragen Sie die WAALTeR-Uhr?* (Ordinale Antwortskala mit fünf Stufen)
- *Wenn Sie die WAALTeR-Uhr tragen, wie lange tragen Sie die Uhr durchschnittlich?* (Ordinale Antwortskala mit sechs Stufen)
- *Wann tragen Sie gewöhnlich die WAALTeR-Uhr?* (Offene Frage)

Dabei ist darauf hinzuweisen, dass die Fragen im Allgemeinen so gestellt wurden, dass sie sich immer auf den *Zeitraum seit dem letzten ausgefüllten Fragebogen* beziehen.

Abbildung 7 stellt die Verteilung der Antwortkategorien in Bezug auf Tragehäufigkeit dar. Hierbei ist mehrgipflige Verteilung der Werte auffällig: Viele TeilnehmerInnen (im Durchschnitt ca. 30%) gaben an, die WAALTeR-Uhr nie zu benutzen. Ebenso gab es einen relativ konstanten Anteil an Personen (im Durchschnitt ca. 20%), die berichteten, die Uhr fast täglich zu benutzen. Dieses Ergebnis weist daraufhin, dass es auf der einen Seite eine distinkte Gruppe von Personen gibt, die die Uhr regelmäßig nutzen und durch sie einen subjektiven Mehrwert erfahren. Auf der anderen Seite gibt es relativ viele Personen, die die Uhr im Prinzip nie verwendet haben.

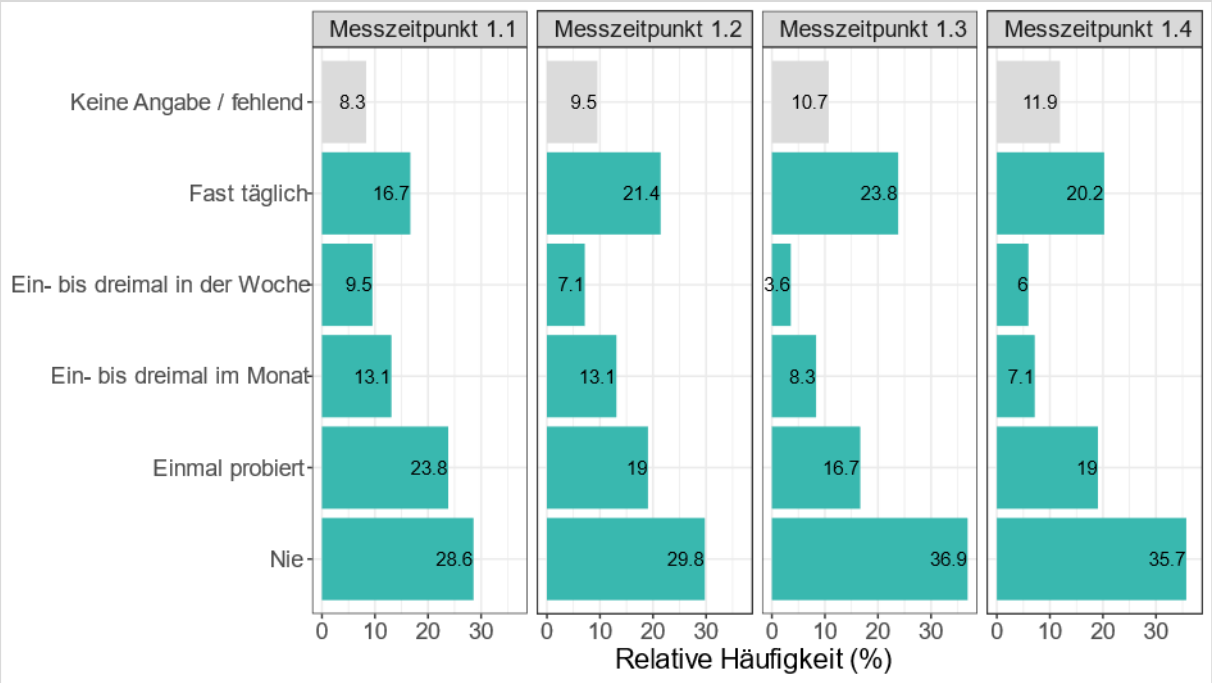

Abbildung 7. Antworthäufigkeiten auf die Frage *„Wie häufig tragen Sie die WAALTeR-Uhr?“* aufgegliedert nach Messzeitpunkten (Studienphase 1).

Abbildung 8 stellt die Antworthäufigkeiten auf die Frage *„Wenn Sie die WAALTeR-Uhr tragen, wie lange tragen Sie die Uhr durchschnittlich?“* dar. Der hohe Anteil an fehlenden Werten ist vermutlich auch durch die hohe Anzahl an Personen bedingt, die die Uhr im Allgemeinen nicht

nutzten und die Frage daher nicht beantworteten. Darüber hinaus ist erkennbar, dass verhältnismäßig viele Personen die Uhr durchschnittlich 4 bis 6 Stunden trugen. Das steht möglicherweise im Zusammenhang mit der Akkulaufzeit der Uhr, die von TeilnehmerInnen häufig bemängelt wurde (siehe auch Abschnitt 5.3.1). Generell ist zu bedenken, dass das Nutzungsverhalten in Bezug auf die Uhr nur anhand von zwei quantitativen Fragen zur Selbsteinschätzung erhoben wurde, die nur eine ungefähre Beurteilung der Nutzungshäufigkeit erlauben.

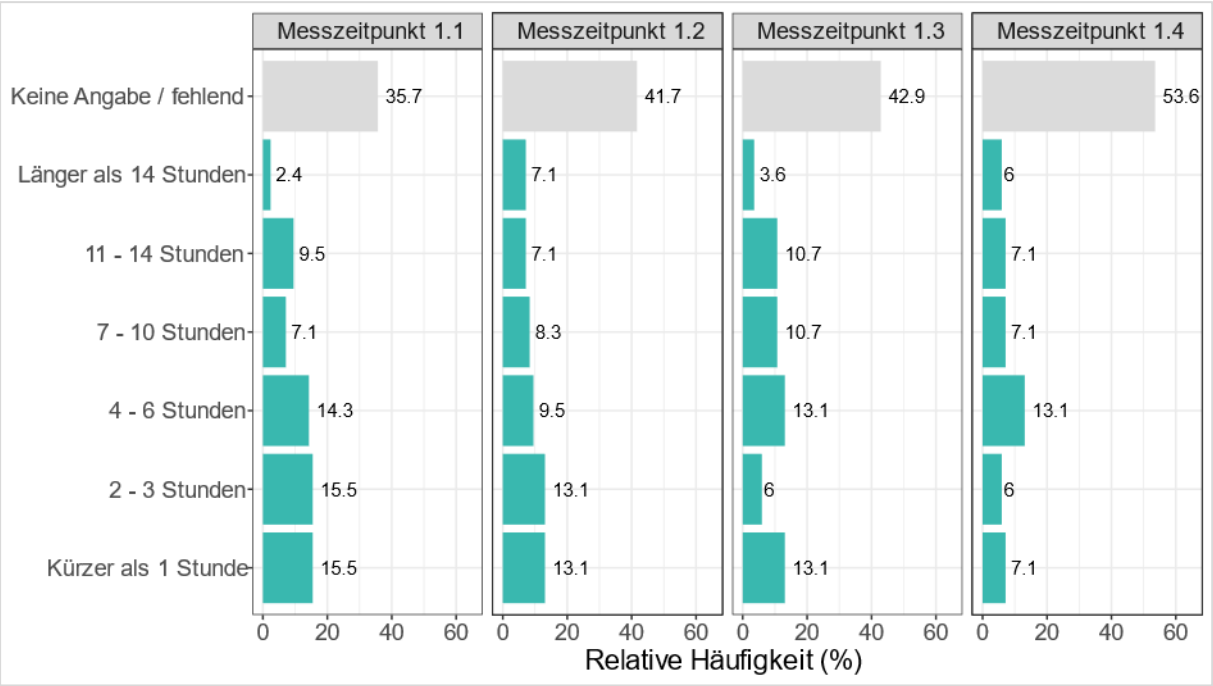

Abbildung 8. Antworthäufigkeiten auf die Frage *„Wenn Sie die WAALTeR-Uhr tragen, wie lange tragen Sie die Uhr durchschnittlich?“* aufgegliedert nach Messzeitpunkten (Studienphase 1).

Abbildung 9. Häufig genannte Begriffe in Zusammenhang mit der Frage *„Wann tragen Sie gewöhnlich die WAALTeR-Uhr?“*. Je häufiger ein Begriff, desto größer ist dessen Darstellung.

Darüber hinaus wurden auf Basis der Frage *„Wann tragen Sie gewöhnlich die WAALTeR-Uhr?“* typische Situationen identifiziert, in denen TeilnehmerInnen die Uhr tragen. Viele TeilnehmerInnen gaben an die Uhr zu tragen, wenn sie (alleine) unterwegs sind. Auch zuhause nutzten nach eigenen Angaben einige TeilnehmerInnen die Uhr. Ein Problem, was in diesem Zusammenhang beschrieben wurde, war die relative schwache Akkulaufzeit der Uhr: Einige Personen berichteten, sie würden die Uhr gerne ständig tragen, dass sei aber aufgrund der geringen Akkulaufzeit nicht möglich. Um die verschiedenen Situationen besser zu veranschaulichen, stellt Abbildung 9 die am häufigsten genannten Begriffe in Zusammenhang mit der Frage *„Wann tragen Sie gewöhnlich die WAALTeR-Uhr?“* grafisch dar.

5.1.3 Bewertung des WAALTeR-Systems

Die Bewertung des WAALTeR-Systems durch die NutzerInnen stellt eine zentrale Variable im Rahmen der Evaluierung dar. Eine positive Bewertung ist wichtige Grundlage für die Nutzung des Systems. Die Nutzung wiederum ist Basis für mögliche Effekte der Nutzung auf Variablen, die Rahmen der Wirkungsanalyse von Interesse sind.

Im Zusammenhang mit der Bewertung eines interaktiven Systems ist das Konstrukt der User Experience von zentraler Bedeutung. User Experience umfasst nach Definition der Internationalen Organisation für Normung (ISO) alle Wahrnehmungen und Reaktionen einer Person, die sich aus der Nutzung sowie der erwarteten Nutzung eines Produkts, Systems oder einer Dienstleistung ergeben (ISO, 2019). Im Rahmen der Evaluierung des WAALTeR-Systems wurde User Experience unter Verwendung des User Experience Questionnaires (UEQ; Laugwitz, Schrepp, & Held, 2006) erhoben. Dieser standardisierte Fragebogen misst User Experience mit 26 Items (semantisches Differential von -3 bis 3) auf folgenden sechs Subdimensionen (Laugwitz, Schrepp, & Held, 2006):

- Attraktivität: Gesamteindruck des Produktes. Gefällt den NutzerInnen das Produkt?
- Durchschaubarkeit: Ist es einfach, sich mit dem Produkt vertraut zu machen? Ist es leicht zu lernen, wie man das Produkt verwendet?
- Effizienz: Können NutzerInnen ihre Aufgaben ohne unnötigen Aufwand lösen?
- Originalität: Ist das Produkt innovativ und kreativ? Weckt das Produkt das Interesse der NutzerInnen?
- Steuerbarkeit: Fühlen sich die NutzerInnen in Kontrolle der Interaktion?
- Stimulation: Ist es spannend und motivierend, das Produkt zu benutzen?

Diese UEQ-Dimensionen wurden zu den Messzeitpunkten 1.1 bis 1.4 mittels Fragebogen erhoben. Die TeilnehmerInnen wurden dazu instruiert ihre Bewertung mittels UEQ hierbei auf das WAALTeR-System als Gesamtsystem (also Tablet und Uhr) zu beziehen. Um Veränderungen über die Zeit in Bezug auf die User Experience inferenzstatistisch zu prüfen, wurden Linear Mixed Effects Models angepasst mit der jeweiligen UEQ-Dimension als abhängiger Variable, Messzeitpunkt als Fixed Effect und Subjekt als Random Effect. Dabei zeigten sich bei keiner der sechs UEQ-Dimensionen signifikante Veränderungen im Studienverlauf, was für eine zeitlich stabile Bewertung des WAALTeR-Systems durch die StudienteilnehmerInnen spricht.

Um eine globale Bewertung des WAALTeR-Systems zu erfassen, die sich einerseits auf die gesamte Studienphase bezieht und andererseits alle Messzeitpunkte gleichermaßen berücksichtigt, wurden die Messungen der UEQ-Dimensionen an den verschiedenen Zeitpunkten zu jeweils einem Mittelwert (pro Dimension, pro TeilnehmerIn) aggregiert. Die aggregierten Skalenmittelwerte wurde in Rahmen der weiteren Auswertung mit Benchmark-Daten, das heißt Vergleichsdaten, in Beziehung gesetzt. Der verwendete Vergleichsdatensatz[8] (Schrepp, Thomaschewski, & Hinderks, 2018) enthält Messungen mittels UEQ von 20190 Personen aus 452 Studien zu verschiedenen Produkten (Unternehmenssoftware, Webseiten, Webshops, soziale Netzwerke). Der Vergleich der Ergebnisse aus der WAALTeR-Evaluierungsstudie mit den Daten im Benchmark erlaubt Rückschlüsse auf die relative Qualität des evaluierten Produkts im Vergleich zu anderen Produkten (Schrepp et al., 2018). Abbildung 10 stellt die aggregierten Skalenmittelwerte in Bezug auf die Referenzbereiche des Vergleichsdatensatzes dar.

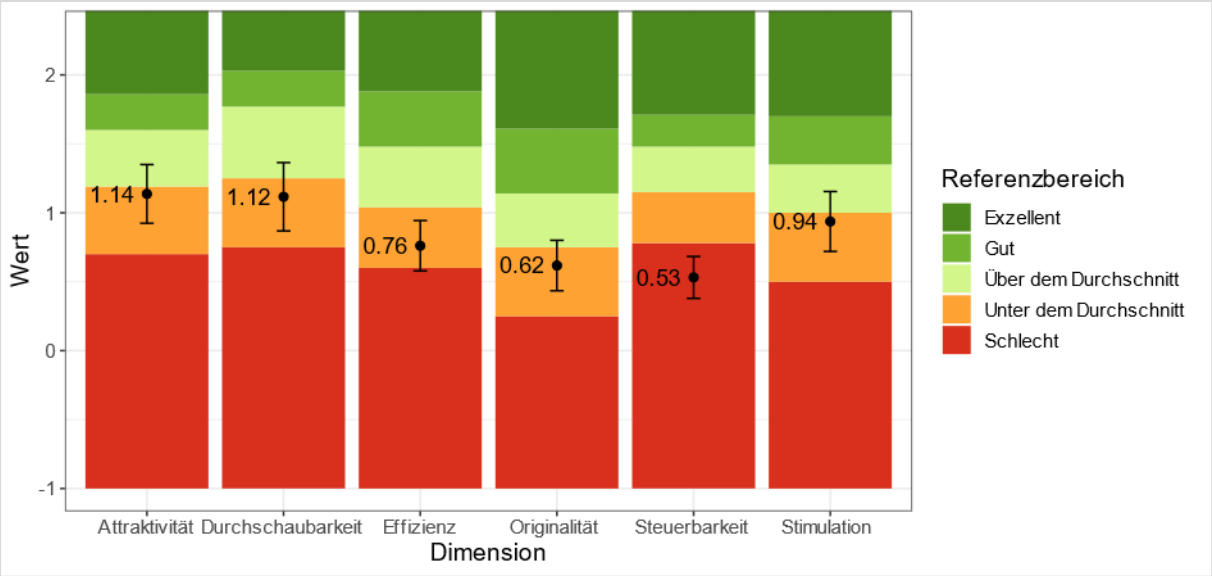

Abbildung 10. Skalenmittelwerte der UEQ-Dimensionen in Bezug auf das WAALTeR-System (Tablet und Uhr) im Vergleich zu Benchmark-Daten nach Schrepp et al. (2018). Vertikale Balken implizieren Konfidenzintervalle (95%). *n* = 150. Wertebereich von -3 bis 3.

Die Referenzbereiche sind nach Schrepp et al. (2018) wie folgt zu interpretieren:

- *Exzellent:* Im Bereich der 10% am besten bewerteten Produkte
- *Gut:* 10% der Produkte wurden besser bewertet, 75% der Produkte schlechter
- *Über dem Durchschnitt:* 25% der Produkte wurden besser bewertet, 50% der Produkte schlechter
- *Unter dem Durchschnitt:* 50% der Produkte wurden besser bewertet, 25% der Produkte schlechter
- *Schlecht:* Im Bereich der 25% am schlechtesten bewerteten Produkte

[8] Verfügbar unter: https://www.ueq-online.org/Material/Data_Analysis_Tools.zip

Grundsätzlich lässt sich erkennen, dass das WAALTeR-System auf allen UEQ-Dimensionen nur unterdurchschnittlich bewertet wurde. Auf der Dimension der Steuerbarkeit liegt das System sogar unterhalb des unteren Quartils (Bereich der 25% am schlechtesten bewerteten Produkte). Diese relativ negativen Bewertungen in Bezug auf das WAALTeR-System hinsichtlich dessen User Experience sind ein potentieller Grund für die geringe Nutzung des Systems sowie ein Indikator für Mängel an der Technologie. In diesem Zusammenhang muss allerdings bedacht werden, dass die herangezogenen Vergleichsdaten in der Regel „allgemeine" Stichproben umfassen (zum Beispiel Stichproben mit Menschen aus vielen unterschiedlichen Altersgruppen), was die Vergleichbarkeit mit den Ergebnissen aus WAALTeR (Stichprobe älterer Menschen einschränken könnte. Diese Annahme kann zum Zeitpunkt der Untersuchung nicht geprüft werden, da eine Überprüfung die Verfügbarkeit normierter Vergleichsdaten (Normierung insbesondere bezüglich des Alters) voraussetzt. Jedenfalls sollten die Ergebnisse aus dem Benchmark-Vergleich mit entsprechendem Bedacht interpretiert werden.

Neben der Messung der User Experience war im Rahmen der Evaluierung von grundlegenden Interesse, ob das WAALTeR-System den Bedürfnissen der TeilnehmerInnen entspricht. Neben qualitativen Fragen (siehe auch Abschnitt 5.3*)* beantworteten TeilnehmerInnen diesbezüglich folgende Frage (Likert-Skala von 1 bis 7):

Wie sehr stimmen Sie folgender Aussage zu: „Das WAALTeR-System entspricht meinen Bedürfnissen nach einem digitalen Assistenz-System."?

Abbildung 11 stellt Antwortverteilungen in Bezug auf diese Frage aufgegliedert nach Messzeitpunkt dar. Ungefähr 30 bis 40% der TeilnehmerInnen stimmen der Aussage zumindest eher zu. Demgegenüber steht allerdings ein erheblicher Anteil an Befragten, deren Bewertung weniger zustimmend ist. Obwohl es sich hier nur um ungefähre Einschätzung anhand einer einzelnen Frage handelt, ist es dennoch ein Anhaltspunkt dafür, dass das WAALTeR-System den Bedürfnissen einer relativ großen Gruppe von TeilnehmerInnen nur unzureichend entspricht. Dieser Aspekt wird im Rahmen der Diskussion erneut aufgegriffen.

Um Gründe für die quantitative Bewertung besser zu verstehen, wurden im Rahmen des Fragebogens auch qualitative Daten zur Bewertung des Systems erhoben. Entsprechend hatten TeilnehmerInnen die Möglichkeit Feedback und Verbesserungsvorschläge zum WAALTeR-System in Form eines offenen Antwortformats zu geben. Auf diese Weise wurden zahlreiche Verbesserungsvorschläge zu einzelnen Anwendungen gesammelt, die Rahmen dieses Berichts nicht im Detail dargestellt werden. Am häufigsten betraf das Feedback jedoch einerseits die WAALTeR-Uhr und andererseits den Wunsch nach zusätzlichen Anwendungen: In Bezug auf die Uhr wurde sehr oft die kurze Laufzeit des Akkus kritisiert. Zusätzliche Anwendungen wurden sowohl für die Uhr als auch für das Tablet gewünscht: Bei der Uhr fehlte vielen TeilnehmerInnen der im Vorfeld angekündigte Schrittzähler, der aufgrund technischer Schwierigkeiten im Rahmen der ersten Studienphase nicht implementiert werden konnte. In Bezug auf das Tablet war die freie Installierbarkeit weiterer Anwendungen ein zentrales Thema: Vergleichsweise viele TeilnehmerInnen störten sich daran, dass keine zusätzlichen Apps auf dem WAALTeR-Tablet installiert werden konnten, so wie es beispielsweise bei Smartphones üblich ist.

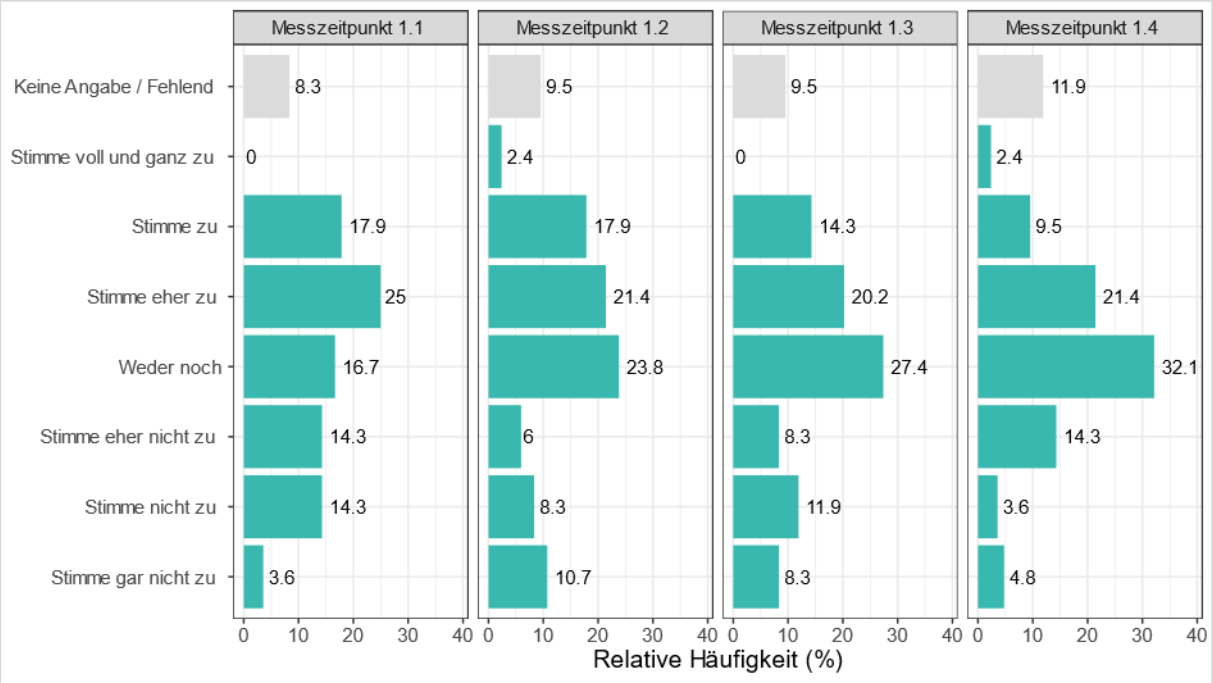

Abbildung 11. „Wie sehr stimmen Sie folgender Aussage zu: «Das WAALTeR-System entspricht meinen Bedürfnissen nach einem digitalen Assistenz-System. »?“. Antwortverteilungen pro Messzeitpunkt.

5.2 Studienphase 2

Aufgrund stark unterschiedlicher Gruppengrößen und der nichtrandomisierten Zuteilung zu Untersuchungsgruppen in der zweiten Studienphase wurde in dieser Studienphase ein starker Fokus auf qualitativen Analysen beziehungsweise die Bewertung der zusätzlichen Technologien durch die TeilnehmerInnen gelegt. Der Blick auf Gruppenunterschiede bezüglich der abhängigen Messgrößen erscheint im Rahmen der zweiten Studienphase problematisch, da die Präferenz für eine bestimmte Technologie mit der jeweiligen abhängigen Variablen zusammenhängen könnte. Beispielsweise wäre es denkbar, dass sich Personen für das Sicherheitspaket entscheiden, weil sie sich im Allgemeinen weniger sicher fühlen. Entsprechend wäre das niedrigere Sicherheitsempfinden in dieser Untersuchungsgruppe offensichtlich nicht durch die Nutzung der Technologie bedingt und die verschiedenen Einflüsse auf die abhängige Variable schlecht voneinander unterscheidbar. Darüber hinaus erschwert das nichtbalancierte Versuchsdesign in dieser Studienphase (stark unterschiedliche Gruppengrößen) die Anpassung von statistischen Modellen mit stabilen Koeffizienten.

5.2.1 Bewertung der zusätzlichen Technologien

5.2.1.1 UEQ-Benchmark-Vergleich

Wie bereits bei Tablet und Uhr wurde in Studienphase 2 die User Experience in Bezug auf die zusätzlichen Technologien (Telemedizinpaket, Sicherheitspaket) unter Verwendung des UEQ erhoben (Laugwitz et al., 2008). Um die relative Qualität der Technologien einzuschätzen, wurden ebenso Benchmark-Daten aus 452 Studien nach Schrepp et al. (2018) herangezogen. Für eine genaue Beschreibung der UEQ-Dimensionen und der Referenzbereiche siehe Abschnitt 5.1.3.

Abbildung 12 stellt die Skalenmittelwerte UEQ-Dimensionen in Bezug auf das Telemedizinpaket im Vergleich zu Benchmark-Daten nach Schrepp et al. (2018) dar. Dabei zeigt sich eine überdurchschnittliche Bewertung der Technologie in Bezug auf die Mehrzahl der UEQ-Dimensionen (Attraktivität, Durschaubarkeit, Originalität, Stimulation). Besonders auffallend ist ein hoher Mittelwert auf der Skala Stimulation. Diese positive Bewertung könnte unter anderem mit dem integrierten Schrittzähler des Telemedizin-Pakets in Zusammenhang stehen. Auch das Sicherheitspaket schneidet im Vergleich mit den Benchmark-Daten gut ab: Auf den Dimensionen Attraktivität und Originalität liegt das System sogar oberhalb des dritten Quartils (Bereich der 25% am besten bewerteten Produkte). Abbildung 13 stellt die entsprechenden Mittelwerte in Bezug auf das Sicherheitspaket dar.

Für beide Produkte liegt Mittelwert bezüglich der Dimension Steuerbarkeit nur im unterdurchschnittlichen Bereich. Diese Dimension beschreibt, inwieweit sich NutzerInnen in Kontrolle der Interaktion fühlen. Es könnte vermutet werden, dass hier auch ein Effekt des Alters bedacht werden muss: Beispielsweise ist vorstellbar, dass ältere Menschen die Steuerbarkeit eines Produkts tendenziell als geringer einschätzen. Die herangezogenen Benchmark-Daten beziehen sich in der Regel auf Stichproben mit größerer Altersspannweite, was die Vergleichbarkeit mit den Daten aus der WAALTeR-Studie einschränken könnte. Ebenso ist zu bedenken, dass vor allem bezüglich des Sicherheitspakets nur eine relativ kleine Stichprobe (n = 11) befragt werden konnte, was die Genauigkeit der Messungen verringert.

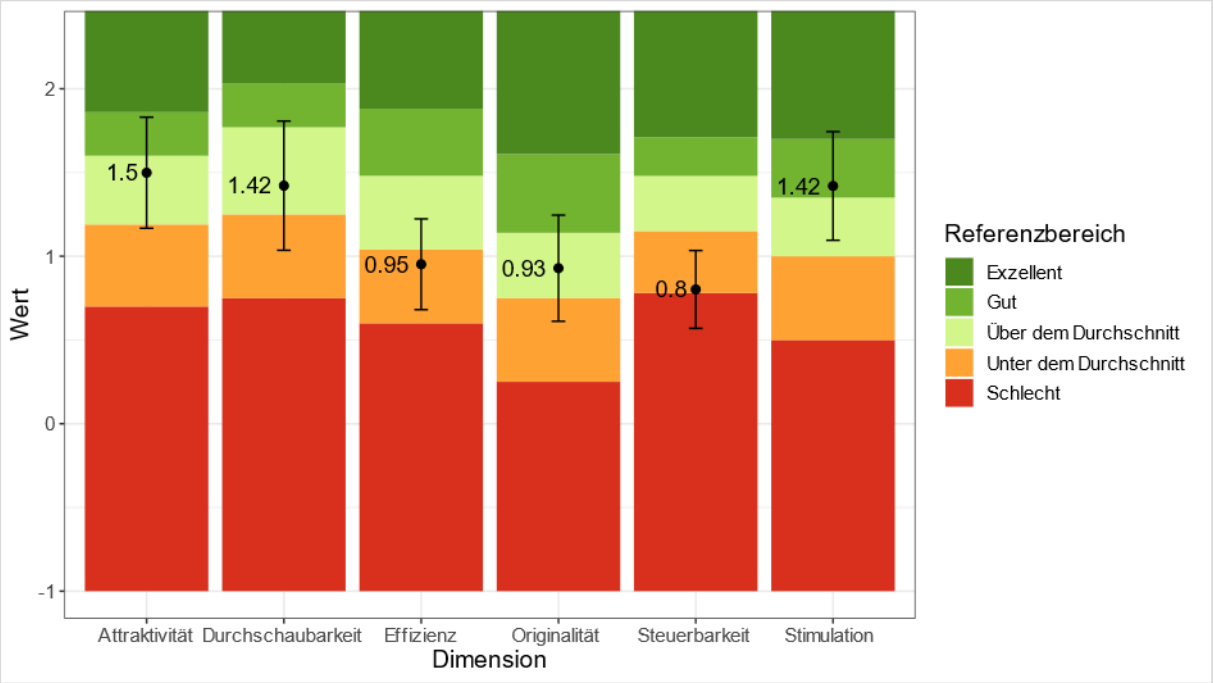

Abbildung 12. Skalenmittelwerte der UEQ-Dimensionen in Bezug auf das Telemedizinpaket im Vergleich zu Benchmark-Daten nach Schrepp et al. (2018). Vertikale Balken implizieren Konfidenzintervalle (95%). *n* = 31. Wertebereich von -3 bis 3.

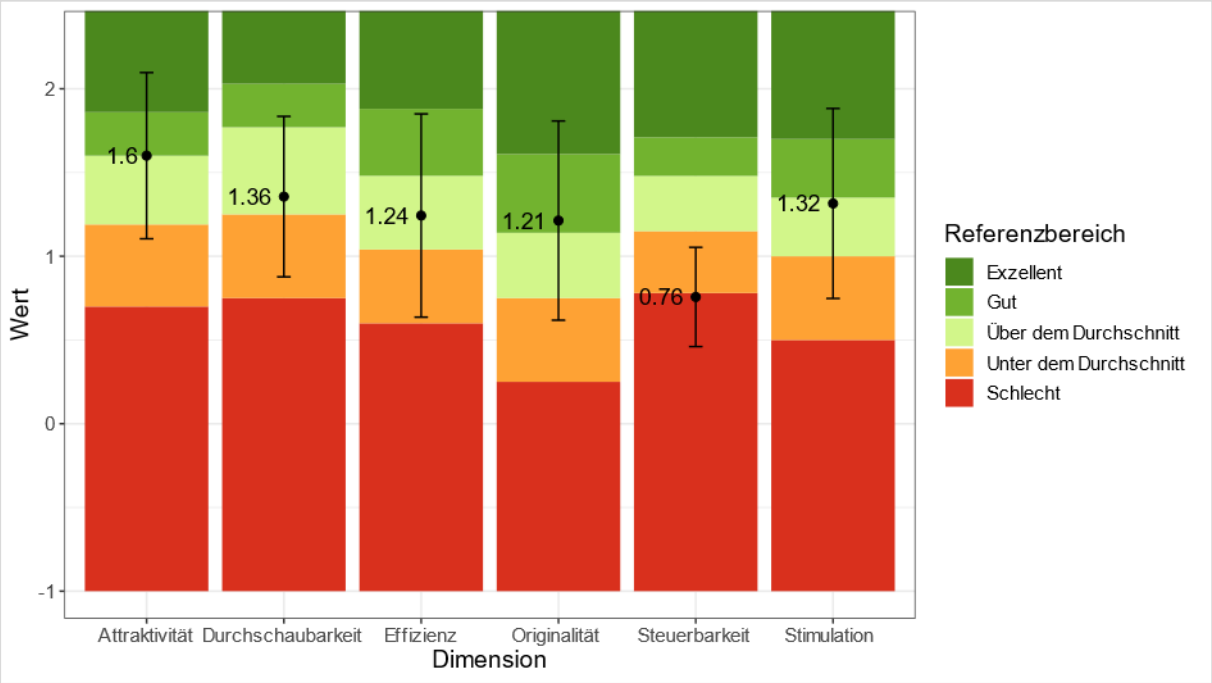

Abbildung 13. Skalenmittelwerte der UEQ-Dimensionen in Bezug auf das Sicherheitspaket im Vergleich zu Benchmark-Daten nach Schrepp et al. (2018). Vertikale Balken implizieren Konfidenzintervalle (95%). *n* = 11. Wertebereich von -3 bis 3.

5.2.1.2 Positive und negative Situationen mit den zusätzlichen Technologien

Neben der quantitativen Erhebung der User Experience wurden die TeilnehmerInnen in der zweiten Studienphase im Rahmen der Fragebögen auch dazu aufgefordert, positive bezeihungsweise negative Situationen mit den Technologien in einem offenen Antwortformat zu beschreiben:

Gab es in den letzten drei Monaten [positive / negative] Situationen dem [Telemedizinpaket / Sicherheitspaket], die Ihnen besonders im Gedächtnis geblieben sind? Falls ja, beschreiben Sie diese kurz.

Die auf diese Weise gesammelten Rückmeldungen wurden themenanalytisch ausgewertet mit dem Ziel, bedeutsame Erfahrungen der TeilnehmerInnen mit der Technologie zu identifizieren. Gerade in Bezug auf das Telemedizinpaket waren die Rückmeldungen positiv. Beispielsweise wurde der Schrittzähler häufig gelobt. Vergleichsweise viele TeilnehmerInnen berichteten, dass sie sich durch den Schrittzähler zu regelmäßiger Bewegung motiviert fühlen:

- *„Diese Sportuhr spornt mich an, noch mehr aktiv zu sein. Jetzt versuche ich auf über 6000 Schritte zu kommen.“*
- *„Schrittzähler motiviert mehr als ich dachte.“*

Außerdem wurde der Aspekt der Vorsorge durch Blutdruckmessung hervorgehoben. TeilnehmerInnen berichteten, dass sie durch die Nutzung des Telemedizinpakets auf die Bedeutung von Blutdruckmessung als gesundheitliche Vorsorgemaßnahme aufmerksam wurden:

- *„Nie hätte ich mich um meinen Blutdruck gekümmert, so gesehen bin ich äußerst dankbar."*
- *„Die Telemedizin ist für mich ein gesundheitliches Plus, da ich sonst nie auf Idee gekommen wäre Blutdruck zu messen."*

Ebenso wurde die Möglichkeit gelobt, das Tablet als digitales Gesundheitstagebuch zu nutzen und Einträge bei Arztbesuchen herzeigen zu können:

- *„Messe täglich Blutdruck und Wohlbefinden. Nehme Tablet immer zum Arzt mit. Super!"*
- *„Jeden Tag dreimal Blutdruckmessen und Schlafphasen-Messung: Sehr hilfreich!"*

Negativ hervorgehoben wurden vorwiegend Probleme im Zusammenhang mit der technischen Zuverlässigkeit des Systems. Beispielweise wurde zurückgemeldet, dass die Kopplung zwischen Messgerät und Tablet unzuverlässig sei. Außerdem wurde die Genauigkeit des Schrittzählers bemängelt:

- *„Aktivitätstracker ausgefallen (einige Tage keine Werte)."*
- *„Verbindung zwischen Tablet und Telemedizin hat nicht funktioniert."*

Bezüglich des Sicherheitspakets waren die Rückmeldungen weniger vielfältig. Jedenfalls wurde die zuverlässige, freundliche Rückmeldung durch das Hausnotrufsystem von den TeilnehmerInnen erwähnt:

- *„Erfahrung, dass „myStella" schnell reagiert."*
- *„Freundliche Begrüßung durch „myStella" beim nach Hause kommen."*
- *„Es war in der Nacht ein Stromausfall. Sofort hat sich jemand über die Stella gemeldet und gefragt, ob alles in Ordnung ist."*

Andererseits wurden häufige Fehlalarme durch den Sturzsensor berichtet. Dabei wurden diverse Gründe für diese Fehlalarme genannt:

- *„Man kann sich nicht länger bücken oder auf den Knien etwas suchen, das löst den Alarm aus."*
- *„Sturzsensor spricht an: Wenn ich etwas aus unterer Lade holen möchte. Wenn man den Bettvorleger gesaugt."*
- *„Fehlalarm als Seitenschläferpolster auf den Boden gefallen ist. Zweiter Fehlalarm bei Öffnen von Schublade."*

Zusammenfassend wurden beide zusätzlichen Technologien von den TeilnehmerInnen tendenziell als nützlich empfunden und unterschiedliche positive Erfahrungen damit berichtet. Negative Kritik bezog sich meist auf die technische Zuverlässigkeit der Anwendungen. Diese auf Basis von Fragebögen identifizieren Themen können offensichtlich nur stichprobenartig einen Eindruck von den Erfahrungen vermitteln, den die TeilnehmerInnen mit den verschiedenen Technologien hatten. Die Ergebnisse aus den qualitativen Interviews (Abschnitt 5.3) vertiefen diese Einblicke und nehmen noch stärkeren Bezug auf die individuellen Erfahrungen der TeilnehmerInnen im Projekt.

5.3 Ergebnisse aus den qualitativen Interviews

5.3.1 Stichprobenbeschreibung

Innerhalb der 18-monatigen Evaluierungsphase wurden auch insgesamt 20 qualitative semi-strukturierte Interviews mit 15 TeilnehmerInnen in zwei Phasen durchgeführt. Nachdem die TeilnehmerInnen das Basis-Paket für ein Jahr benutzt hatten (Messzeitpunkt 1.4), wurden 15 semi-strukturierte Interviews durchgeführt, welche jeweils etwa eine Stunde dauerten. Nach den Interviews erhielten die TeilnehmerInnen ein kleines Geschenk als Aufwandsentschädigung. Die TeilnehmerInnen wurden auf Basis von Diversitätskriterien ausgewählt, wobei Gemeindebezirk, Gender, technische Affinität, Alter, beruflicher Hintergrund der TeilnehmerInnen und Haushaltsgröße (d.h. Anzahl der Personen im Haushalt) berücksichtigt wurden. Gegen Ende der Evaluierungsstudie (Messzeitpunkt 2.2) wurden 5 weitere semi-strukturierte Interviews mit Personen durchgeführt, welche mit zusätzlichen Paketen ausgestattet wurden. Diese dauerten je etwa 30 Minuten. Die TeilnehmerInnen konnten den Ort der Interviews wählen. Alle Interviews wurden transkribiert und mittels thematischer Analyse nach Braun und Clarke (2006) analysiert.

5.3.2 Ergebnisse

Die Motivationen und Ziele der TeilnehmerInnen innerhalb des WAALTeR-Projekts waren sehr divers. Einige sprachen von Neugierde und wollten sich über Möglichkeiten informieren, *„abseits von Pflegeheimen“* zu altern und länger in den eigenen vier Wänden zu wohnen. Andere TeilnehmerInnen sprachen darüber, mit ihrer eigenen Expertise und ihrem Engagement zum Forschungsprojekt beizutragen. Darunter waren auch TeilnehmerInnen, die entweder ehemals oder nach wie vor beruflich mit medizinischen Technologien zu tun hatten oder im Gesundheitsbereich tätig waren. Andere TeilnehmerInnen waren motiviert, die Technologien kennenzulernen oder ihre eigenen digitalen Kompetenzen zu verbessern. Zudem waren TeilnehmerInnen daran interessiert, sich mit den Technologien um andere Menschen zu kümmern, z.B. um Partner oder im Rahmen von freiwilliger Arbeit. Darüber hinaus gab es TeilnehmerInnen, die überwiegend soziale Interessen verfolgten und z.B. an den Stammtischen interessiert waren, wo es möglich war, andere Personen sowie deren Umgang mit Technologien kennenzulernen („*Es geht wirklich nur um die sozialen Kontakte*“).

Die soziale Infrastruktur des Projekts, d.h. vor allem die Stammtische, erwiesen sich als wesentlich für viele TeilnehmerInnen. Dabei waren viele daran interessiert, andere TeilnehmerInnen zu treffen. Für andere kamen die Stammtische ihrem Interesse entgegen, ihre eigenen digitalen Kompetenzen weiterzuentwickeln. Manche TeilnehmerInnen waren bereits vor der Teilnahme am Projekt befreundet gewesen, andere waren daran interessiert, neue Kontakte zu knüpfen. Im Gegensatz zu den Stammtischen wurden die bereitgestellten Apps, welche eine soziale Vernetzung ermöglichen (wie z.B. die Anwendung zur Nachbarschaftsvernetzung) allerdings abgelehnt. Die TeilnehmerInnen begründeten den fehlenden Nutzen damit, dass sie keine Zeit hatten oder bereits sozial gut vernetzt waren, auch in ihrer Nachbarschaft. Zudem waren einige TeilnehmerInnen skeptisch, zu viele Informationen an ihr näheres Umfeld preiszugeben, da sie sich vor Einbrüchen fürchteten.

Die gefühlte Sicherheit spielte im Projekt ebenfalls eine Rolle. Beispielsweise wurde die Uhr von einigen TeilnehmerInnen als vorteilhaft empfunden, auch, da sie nicht nur in den eigenen vier Wänden verwendet werden könnte. Dabei wäre es aber immer wieder wichtig, „*die Hunde wach* [zu] *halten",* sodass der Bereitschaftsdienst im tatsächlichen Notfall auch zur Verfügung stünde. Andererseits waren TeilnehmerInnen auch aufgrund technischer Probleme der Uhr skeptisch, bevor diese ausgetauscht wurde.

Die Technologien und der Umgang damit wurden unterschiedlich bewertet. Einige TeilnehmerInnen kritisierten das mangelnde Angebot („*das ist Vorschule*"), wobei andere Schwierigkeiten hatten, das volle Potential der Anwendungen auszuschöpfen, obwohl sie das gerne getan hätten. Vor allem die Anwendungen am Tablet wurden dafür kritisiert, dass sie Hürden boten, wie z.B. nicht nachvollziehbare Passwortanfragen oder Navigationshürden etwa im Veranstaltungskalender. Die Stammtische und auch die Interviews wurden auch dafür genützt, um Hilfe im Umgang mit den Anwendungen zu erhalten. Allerdings hatten einige TeilnehmerInnen auch das Gefühl, sie wären die einzigen, die sich wenig mit den Technologien auskennen oder sie suchten die Schuld bei sich selbst („*vielleicht bin ich nicht so talentiert?*").

Die Interviews ergaben zudem einen verstärkten Bedarf an Personalisierung vor allem in Hinblick auf die Technologien, aber zum Teil auch bei der Unterstützung im Umgang damit. Vor allem Personen mit weniger digitalen Kompetenzen gaben an, sich mehr Unterstützung zu wünschen („*ich brauche eine persönlichere Einschulung*"). Für das Angebot zu Veranstaltungen oder Online-Medien herrschte Interesse und Nutzungsbereitschaft, jedoch fehlte häufig der Bezug zur eigenen Lebenswelt. Etwa waren die Angebote der E-Zeitungen nur auf einige Herausgeber beschränkt, oder das Veranstaltungsangebot verfügte nicht über die bekannten Veranstaltungsorte. Daher bevorzugten die TeilnehmerInnen das entsprechende Angebot in Print-Medien, oder sie nutzten andere Apps („*ich schau mir also an was, was wird von Veranstaltungen angeboten, dann les ich irgendwann auch einen Artikel von einem Konkurrenzblatt*"). Trotz bekundetem Interesse an Veranstaltungen verstanden sich die TeilnehmerInnen nicht immer als Zielgruppe der konkreten Angebote („*Da, Geburtsvorbereitung für werdende Eltern*[, …] *ich bin jetzt Uroma von einem*"). Dies wurde in den Interviews zum Teil mit Humor genommen, führte jedoch auch zu einer Infragestellung der eigenen Gewohnheiten („*wer geht denn am Freitag am Abend schon in die Josefstadt?*"). Die TeilnehmerInnen erzählten auch von anfänglichem Interesse an dem Kartensystem, wobei die vorgeschlagenen Wege jedoch keine Verbesserung zur eigenen Orientierung darstellten („*da hab ich mir halt gedacht vielleicht kann mir das Gerät einen geschickteren Weg sagen als ich weiß, hat mir aber auch nichts gescheiteres sagen können also dann war das Radln noch das allerbeste*").

Auch auf der Design-Ebene gab es manchmal Probleme mit der konkreten Ausgestaltung. Tablets an sich erwiesen sich als praktisch für zu Hause, jedoch für einige TeilnehmerInnen auch als zu schwer für unterwegs oder unhandlich, was etwa das folgende Zitat einer gesunden, älteren Teilnehmerin veranschaulicht: „*Es ist ein bisschen unhandlich groß, ich hab so glatte Hände und da muss ich man sehr aufpassen, dass es nicht aus der Hand rutscht*". Die konkrete Teilnehmerin schlug Einkerbungen vor, um das Gerät besser in der Hand halten zu können.

Unter den interviewten Personen, die Blutdruckmessgeräte erhielten, wurden diese begrüßt, auch, da sich das tägliche Messen des Blutdrucks gut in den eigenen Lebensalltag integrieren ließ („*das ist wie Zähne putzen*“). Der Sturzsensor musste zwar gelegentlich ausgeschaltet werden, um die fälschliche Erkennung von sich auf dem Boden befindenden oder bewegenden Objekten (z.B. Staubsauger) zu vermeiden. Andererseits wurde der Sturzsensor grundsätzlich als sinnvoll empfunden, wobei auch Überlegungen dahingehend getätigt wurden, inwiefern eine Anbringung in einem einzigen Zimmer ausreichend wäre. Die interviewten TeilnehmerInnen gaben auch an, dass sich der Sensor ästhetisch in ihre Wohnung integrieren ließ.

Bedenken wegen der Privatsphäre wurden allgemein geäußert, wobei diese in Zusammenhang mit der eigenen digitalen Kompetenz bzw. Wissen gebracht wurde (z.B. „*eine ganze Menge an Menschen haben Befürchtungen* […] *weil sie sich nicht auskennen*“). Jedoch brachten die Teilnehmerinnen auch zum Ausdruck, dass innerhalb des Projekt-Kontexts grundsätzlich Vertrauen vermittelt würde, und zwar auch in Hinblick auf persönliche Daten.

5.4 Weitere Analysen

5.4.1 Zufriedenheit mit der Studie

Zusätzlich wurden TeilnehmerInnen im Rahmen der Fragebögen zu ihrer allgemeinen Zufriedenheit mit der WAALTeR-Studie befragt (Likert-Skala von 1 bis 7, gemeinsam mit Freitextantworten, in der die TeilnehmerInnen die Möglichkeit ihre Bewertung begründen). Diese Erhebung umfasst also nicht nur die Zufriedenheit mit der Technologie, sondern die Zufriedenheit mit der Studienteilnahme insgesamt. Dies erlaubt Rückschlüsse auf die Motivation zur Teilnahme und den persönlichen Mehrwert, den die Studie für die TeilnehmerInnen besitzt.

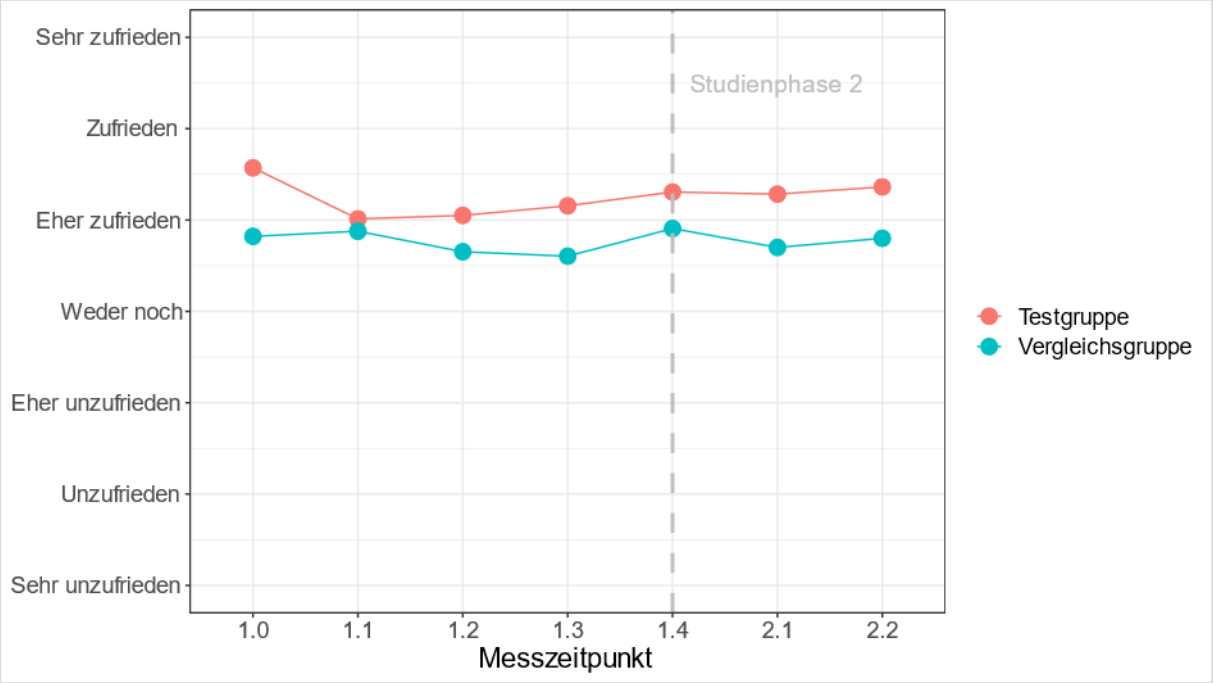

Abbildung 14. Mittelwerte der Zufriedenheit mit der Studienteilnahme nach Messzeitpunkt und Untersuchungsgruppe.

Wie in Abbildung 14 zu erkennen, zeigen sich beide Gruppen tendenziell zufrieden mit der Studienteilnahme. Im Gruppenvergleich beschreibt sich die Testgruppe tendenziell als etwas zufriedener als die Vergleichsgruppe. Ein häufig genannter Grund für die Zufriedenheit mit der Teilnahme war in beiden Gruppen die gute Betreuung während des Projekts:

- *„Das Projekt entspricht meinen Erwartungen. Das Angebot ist nützlich, die Betreuung hervorragend. Sehe der weiteren Durchführung mit Interesse entgegen*."

Außerdem wurde von vielen TeilnehmerInnen beider Gruppen betont, dass sie das Gefühl haben, durch ihre Teilnahme etwas beizutragen:

- *„Ich halte diese Studie für wichtig. Ich bin froh, dazu einen Beitrag zu leisten."*
- *„Ich denke, ich konnte was beitragen!"*

Von Personen der Testgruppe wurde oft positiv hervorgehoben, dass durch die Teilnahme am Projekt, ein Heranführen an neue Technologien stattgefunden hat, was für die TeilnehmerInnen persönlichen Mehrwert brachte:

- *„Ich bin zufrieden, weil ich viele neue Ideen gewonnen habe."*
- *„Ich lerne Möglichkeiten der Unterstützung kennen, auch wenn ich sie derzeit noch nicht brauche."*
- *„Die Beschäftigung mit dem Tablet hat mich ermutigt, mir ein Smartphone anzuschaffen."*

Allerdings wurde von den TeilnehmerInnen auch häufig erwähnt, dass sie sich noch mehr technische Einschulung gewünscht hätten:

- *„Ich vermisse mehr technische Einschulung."*
- *„Zur Zufriedenheit fehlt mir eine gründliche Einschulung, um alle Funktionen des Systems in Anspruch nehmen zu können."*

Darüber hinaus waren fehlende oder fehlerhafte Funktionalitäten ein oft genannter Grund für die Unzufriedenheit der Teilnahme der Testgruppe, (z.B., wenn bestimmte Anwendungen nicht richtig funktionierten beziehungsweise fehlten):

- *„Im Vergleich zu PC / Handy gibt es keine zusätzlichen Anwendungen, die besser sind."*
- *„Wegen der Nichtfunktion der Uhr ist eine anfängliche Begeisterung gedämpft."*

Seitens der Vergleichsgruppe wurde relativ häufig erwähnt, dass man als TeilnehmerIn der Vergleichsgruppe weniger Einblick in das Projekt hatte und dass man sich durch die Teilnahme auch Technologie (Tablet und Uhr) erhofft hatte:

- *„Ich wäre lieber in der Testgruppe gewesen."*
- *„Ich gehöre zur Vergleichsgruppe, bin daher nicht so nah dabei."*

Insgesamt zeigen diese qualitativen Rückmeldungen zur Teilnahmezufriedenheit, dass die TeilnehmerInnen vor allem am Kennenlernen und der Auseinandersetzung mit neuer Technologie (Smartphone, Tablet, etc.) interessiert sind. Dabei wird die Heranführung beziehungsweise Einschulung in Technologie als besonders wichtig empfunden und stellt einen Mehrwert für die Personen dar.

5.5 Zusammenfassung der Ergebnisse in Bezug auf die Fragestellungen

Fragestellung 1: Beeinflusst die Nutzung des WAALTeR-Systems die Selbstbestimmtheit, Autonomie, Soziale Partizipation und das Sicherheitsempfinden älterer Menschen?

Zusammenfassend zeigten sich auf Basis des experimentellen Untersuchungsdesigns keine signifikanten Effekte der Technologie auf die untersuchten Messgrößen Selbstbestimmtheit, Autonomie, Soziale Partizipation und Sicherheitsempfinden. Dies trifft sowohl auf Veränderungen über Zeit als auch auf Unterschiede zwischen Gruppen zu. Beim Vergleich der Untersuchungsgruppen zum abschließenden Messzeitpunkt der ersten Studienphase weisen die Daten auf einen kleinen Effekt bezüglich Sicherheitsempfinden und der BMPN-Subskala Autonomy hin. Die Richtung der Effekte ist dabei erwartungskonform: Personen der Testgruppe berichteten zum abschließenden Messzeitpunkt höheren Werte in Bezug auf beide Messungen. Beide Effekte erreichten allerdings keine Signifikanz. Das könnte auch in Zusammenhang mit der statistischen Power der Analyse stehen.

Fragestellung 2: Wie wurde das WAALTeR-System genutzt (Häufigkeit, Dauer, Anwendungen)?

In Bezug auf das WAALTeR-Tablet implizieren die erhobenen Nutzungsdaten, dass das Tablet in der Regel nicht tagtäglich Verwendung fand, sondern eher gelegentlich in die Hand genommen wurde, das heißt im Durchschnitt an ein bis drei Tagen pro Woche. Insgesamt am häufigsten genutzt wurden Anwendungen wie der Webbrowser, E-Mail, WhatsApp, die Zeitungs-App und verschiedene Spiele. Auch zeigt sich, dass die Nutzung des Tablets sehr heterogen war: Die individuellen NutzerInnen haben die verschiedenen Anwendungen jeweils unterschiedlich häufig verwendet. Bezüglich der WAALTeR-Uhr weisen die Daten darauf hin, dass eine distinkte Gruppe an TeilnehmerInnen (ca. 20%) die Uhr fast täglich trug.

Fragestellung 3: Wie wird das System hinsichtlich seiner User Experience von den NutzerInnen bewertet?

In Referenz zu Benchmark-Daten wurde die User Experience des WAALTeR-Systems (Tablet und Uhr) nur unterdurchschnittlich bewertet. Das weist auf Mängel bezüglich des Nutzungserlebnisses hin. Dieses Ergebnis sollte jedoch mit großem Bedacht interpretiert werden, da die herangezogenen Benchmark-Daten in der Regel auf Stichproben mit größerer Altersspannweite beruhen. Die zusätzlichen Technologien in der zweiten Studienphase (Telemedizin-Paket und Sicherheitspaket) wurden hinsichtlich ihrer User Experience auf Basis quantitativer Daten positiv bewertet. Darüber hinaus lassen qualitative Rückmeldungen vor allem in Bezug auf das Telemedizinpaket auf eine große Zufriedenheit der TeilnehmerInnen mit der Technologie schließen.

Fragestellung 4: Welche Rolle haben Technologien wie WAALTeR in der Lebenswelt älterer Menschen?

Die Zielgruppe älterer Menschen erscheint heterogen mit verschiedenen individuellen Bedürfnissen hinsichtlich der Nutzung von Technologie. Das zeigt sich anhand der Nutzungsdaten und auf Grundlage der qualitativen Interviews. Diese Bedürfnisse hinsichtlich Technologie stehen offensichtlich in Wechselwirkung mit den Lebenswelten der TeilnehmerInnen. In diesem Kontext wurde auch der Wunsch nach stärkerer Personalisierbarkeit der Technologie durch die TeilnehmerInnen hervorgehoben. Dabei ist zu bedenken, dass bereits kleine Unstimmigkeiten bei der Verwendung dazu führen können, dass TeilnehmerInnen keinen Nutzen in der Technologie gesehen haben. Zentrale Motive, die Personen zur Teilnahme am WAALTeR-Projekt bewegten, sind das Kennenlernen neuer Technologie und die soziale Komponente des Projekts. Das Bedürfnis nach der Einschulung in Technologie ist dabei in hohem Maß vorhanden.

6 Diskussion

Die im Rahmen von WAALTeR durchgeführte Studie evaluierte eine tabletbasierte AAL-Technologie anhand eines experimentellen Untersuchungsdesigns mit über 140 TeilnehmerInnen. Mit dieser methodischen Herangehensweise adressiert sie eine wesentliche Forschungslücke, denn die Anzahl an experimentellen Designs zur Evaluierung von AAL- und vergleichbaren Technologien ist in der bisherigen Forschung vergleichsweise gering. Die Daten der Wirkungsanalyse konnten keine signifikanten Effekte der Technologie auf die erhobenen Messgrößen zeigen. Dennoch liefert das Projekt auf Basis der umfassenden qualitativen und quantitativen Daten viele bedeutsame Erkenntnisse sowohl für die künftige Forschung als auch für Initiativen mit dem Ziel digitale Inklusion älterer Menschen zu fördern.

6.1 Reflexion und Empfehlungen für zukünftige Studien

Im folgenden Abschnitt werden mögliche Gründe für die vorliegenden Ergebnisse aus der Wirkungsanalyse exploriert. Diese Gründe ergeben sich einerseits aus der Methodik der Untersuchung und andererseits aus Merkmalen der evaluierten Technologie selbst. Des Weiteren werden auf Basis der Ergebnisse der WAALTeR-Evaluationsstudie mögliche Implikationen für zukünftige Forschung im AAL-Bereich diskutiert.

Wie bereits erwähnt wurden keine signifikanten Effekte durch die Nutzung von WAALTeR in den gewählten Messgrößen gefunden. Aspekte der Methodik, die dazu führen könnten, dass tatsächlich vorhandene Effekte nicht gefunden werden, könnten in Charakteristika der Stichprobe, der Art der Messungen und in der statistischen Power der Analyse liegen. Die untersuchte Stichprobe weist einige Merkmale auf, die die Feststellbarkeit von Effekten erschweren und die Generalisierbarkeit von Ergebnissen verringern könnten. So deuten die Verteilungen der abhängigen Variablen in der untersuchten Stichprobe auf Deckeneffekte hin. Das wird insbesondere beim Vergleich mit Normstichproben deutlich: Zum Beispiel berichten Conrad et al. (2016) für die WHOQOL-OLD-Facetten Autonomie und Soziale Partizipation Stichprobenmittelwerte von je 68.49 (SD = 19.05) und 69.08 (SD = 20.10) in Bezug auf eine Stichprobe aus der deutschen Allgemeinbevölkerung im Alter von 60 Jahren und älter. Die Gesamtstichprobe im Rahmen der WAALTeR-Evaluierungsstudie berichtete hingegen Mittelwerte von jeweils 74.78 (SD = 13.14) und 74.28 (SD = 15.73) für Autonomie und Soziale Partizipation. Das weist daraufhin, dass die in WAALTeR untersuchte Stichprobe aus sich heraus bereits eine vergleichsweise hohe Lebensqualität besitzt und eine Technologie wie WAALTeR dadurch nur geringe Wirkung haben kann. Darüber hinaus handelt es sich bei den untersuchten Personen um eine selbstselektierte Stichprobe. Das zeigt sich unter anderem in der Geschlechterverteilung mit einem Frauenanteil von ungefähr 75% in der untersuchten Stichprobe. Offensichtlich entspricht dieser Anteil nicht der Geschlechterverteilung in der österreichischen Referenzpopulation (Statistik Austria, 2019). Obwohl solche Unterschiede aufgrund praktischer Überlegungen hinsichtlich Stichproberekrutierung oft in Kauf genommen werden müssen, schränken sie die Generalisierbarkeit der Ergebnisse ein.

Außerdem könnten die verwendeten Messungen zu unspezifisch sein, um mögliche Einflüsse der Technologie aufzudecken. Zwar wurden nach Möglichkeit normierte, psychometrisch validierte Skalen verwendet, doch diese beziehen sich auf umfassende psychologische Konstrukte, die potentiell durch viele verschiedene Faktoren beeinflusst werden können. Daher ist auch die Weiterentwicklung von geeigneten Messinstrumenten eine wichtige Aufgabe zukünftiger Forschung. Einen bedeutsamen Beitrag dazu liefert beispielsweise das Projekt EvAALuation[2] (Himmelsbach et al., 2019).

Darüber hinaus scheint das verwendete Untersuchungs- beziehungsweise Analysedesign auf Basis von Poweranalysen keine ausreichende statistische Power zu besitzen, um kleine Effekte zu detektieren.[9] Das bedeutet nicht, dass die Effekte in WAALTeR mit einer größeren Stichprobe jedenfalls signifikant gewesen wären. Stattdessen ist diese Beobachtung vor allem für die Planung zukünftiger Untersuchungen bedeutsam. Die geringen Effektstärken, die in der WAALTeR-Evaluierungsstudie gefunden wurden, stehen im Gegensatz zu der Studie von Dubuy et al. (2016), die auch mittlere bis große Effektstärken berichtete. Das legt den Schluss nahe, dass – je nach spezifischer Technologie – die Effektstärken in Bezug auf die Wirkung von Technologie unterschiedlich groß sind. Bei der Planung künftiger Studien kann diese Beobachtung miteinbezogen werden – insbesondere in Bezug auf die Fallzahlplanung.

In Bezug auf die Technologie selbst weisen die erhobenen Daten auf einige Mängel im Nutzungserlebnis hin. Zudem wurde das WAALTeR-System individuell unterschiedlich genutzt: Das Tablet umfasste eine Vielzahl verschiedener Anwendungen, die von den TeilnehmerInnen jeweils unterschiedlich oft verwendet wurden. Dadurch könnten eindeutige Effekte der Technologie nur schwer zu finden sein. Auch die Personalisierbarkeit ist in diesem Kontext zu beachten, da sie prinzipiell die Herstellung einer besseren Passung zwischen individuellen Bedürfnissen und Technologie ermöglichen kann. In der Studie von Dubuy et al. (2016) stand diese Personalisierbarkeit von Technik im Mittelpunkt des Designprozesses. Jener Aspekt war in Bezug auf das WAALTeR-System weniger stark ausgeprägt - zum Beispiel konnten keine zusätzlichen Apps installiert werden, wie es etwa bei Smartphones möglich ist. TeilnehmerInnen in WAALTeR äußerten entsprechend oft den Wunsch nach mehr Möglichkeiten zur Anpassung der Technologie. Wir empfehlen daher die Personalisierbarkeit von Technologie als wichtigen Grundsatz im Rahmen zukünftiger Forschung und Entwicklung im AAL-Bereich zu berücksichtigen.

Letztendlich ist eine Evaluierungsstudie zur Wirkung von Technologie nur teilweise vergleichbar mit einer klinischen Interventionsstudie. Es kann davon ausgegangen werden, dass der Effekt von Technologie als Intervention stärker von äußeren Faktoren beeinflusst wird als der Effekt eines Medikaments – auch weil der Anwendungskontext (welche App wurde verwendet?) und Anwendungshäufigkeit je nach StudienteilnehmerIn variieren kann. Wie WAALTeR zeigt, ist die Nutzung von Technologie außerdem stets eingebettet in ein umgebendes soziotechnisches System und kleine Unstimmigkeiten bei der Verwendung können dazu führen, dass kein Nutzen in der Technologie gesehen wird. Das ist eine inhärente Herausforderung

[9]Statistische Power bezeichnet die Wahrscheinlichkeit, dass ein Effekt entdeckt wird, wenn ein Effekt auch tatsächlich existiert.

bei der Evaluierung von Technologie, die eine überlegte Abwägung bezüglich der Auswahl von Untersuchungsmethoden verlangt. Je nach Rahmenbedingungen (z.B. Art der zu evaluierenden Technologie, Ressourcen, Art der Stichprobe, Verfügbarkeit sinnvoller Messungen) ist ein Fokus auf qualitative Methoden möglicherweise die sinnvollste Wahl. Experimentelle Untersuchungen in diesem Kontext verlieren dadurch nicht ihren Wert – allerdings ist zu deren sinnhaftem Einsatz eine umfassende methodenbezogene Reflexion gefordert.

6.2 Konklusion

Die Evaluierung von Technologie bringt eine Vielzahl von Herausforderungen mit sich. Unter anderem könnten Effekte von Technologien wie WAALTeR kleiner sein als bisher angenommen, was bei der Planung zukünftiger Studien berücksichtigt werden kann. Qualitative Methoden können einen wichtigen Beitrag leisten, um quantitative Ergebnisse zu ergänzen und den individuellen Anwendungskontext besser zu verstehen.

Darüber hinaus zeigt sich auf Basis der qualitativen Daten, dass sowohl die Technologie selbst als auch die Teilnahme an einem Projekt wie WAALTeR älteren Menschen Mehrwert bringt. Dabei sind das Kennenlernen von neuer Technologie und der soziale Aspekt des Projekts (Stammtische, Veranstaltungen) wichtige Motive, die Personen zur Teilnahme bewegten und zu ihrer Zufriedenheit mit der Studienteilnahme beitrugen. Somit liegt der Mehrwert des Projekts nicht nur auf einer wissenschaftlichen, sondern auch auf einer sozialen und bildungsbezogenen Ebene: Projekte wie WAALTeR können auch dazu beitragen, ältere Menschen an neue Technologien heranzuführen und so digitale Kompetenzen in dieser Altersgruppe zu fördern.

7 Literaturverzeichnis

Bache, S. M., & Wickham, H. (2014). magrittr: A Forward-Pipe Operator for R. Abgerufen von: https://CRAN.R-project.org/package=magrittr

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics, 29(4), 1165–1188. https://doi.org/10.1214/aos/1013699998

Benoit, K., Muhr, D., & Watanabe, K. (2019). stopwords: Multilingual Stopword Lists. Abgerufen von: https://CRAN.R-project.org/package=stopwords

Bortz, J., & Döring, N (2006). Forschungsmethoden und Evaluation (4. Auflage). Heidelberg: Springer.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. doi: 10.1191/1478088706qp063oa

Champely, S. (2018). pwr: Basic Functions for Power Analysis. Abgerufen von: https://CRAN.R-project.org/package=pwr

Cheng, J., Karambelkar, B., & Xie, Y. (2019). leaflet: Create Interactive Web Maps with the JavaScript „Leaflet“ Library. Abgerufen von: https://CRAN.R-project.org/package=leaflet

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale: Erlbaum.

Conrad, I., Matschinger, H., Kilian, R., & Riedel-Heller, S. (2016). WHOQOL-OLD und WHOQOL-BREF – Handbuch für die deutschsprachigen Versionen der WHO-Instrumente zur Erfassung der Lebensqualität im Alter. Göttingen: Hogrefe.

Dubuy, L., Conel, C., & Sauzéon, H. (2016). Self determination-based design to achieve acceptance of assisted living technologies for older adults. Computers in Human Behaviour, 65, 508 – 521. doi:10.1016/j.chb.2016.07.042

Fai, A.H., Cornelius, P.L. (1996). Approximate F-Tests of Multiple Degree of Freedom Hypotheses in Generalised Least Squares Analyses of Unbalanced Split-Plot Experiments. Journal of Statistical Computation and Simulation, 54(4), 363–378. doi:10.1080/00949659608811740.

Fellows, I. (2018). wordcloud: Word Clouds. Abgerufen von: https://CRAN.R-project.org/package=wordcloud

Garnier, S. (2018). viridis: Default Color Maps from 'matplotlib'. Abgerufen von: https://CRAN.R-project.org/package=viridisLite

Giesbrecht F.G., & Burns J.C., (1985). Two-Stage Analysis Based on a Mixed Model: Large Sample Asymptotic Theory and Small-Sample Simulation Results. Biometrics, 41(2), 477–486. doi:10.2307/2530872.

Green P., & MacLeod C.J. (2016). simr: An R package for power analysis of generalised linear mixed models by simulation. Methods in Ecology and Evolution, 7(4), 493-498. doi: 10.1111/2041-210X.12504.

Grolemund, G., & Wickham, H. (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1–25.

Henry, L., & Wickham, H. (2019). purrr: Functional Programming Tools. Abgerufen von: https://CRAN.R-project.org/package=purrr

Hester, J., & Wickham, H. (2019). fs: Cross-Platform File System Operations Based on „libuv“. Abgerufen von: https://CRAN.R-project.org/package=fs

Himmelsbach, J., Gerdenitsch, C., Schwarz, S., Neureiter, K., Garschall, M., Birgit Aigner-Walder, … Oberzauche, J. (2019). EvAALuation[2]. Handbuch für Messinstrumente zur Bewertung von AAL Lösungen. Wien, Villach. Abgerufen von: https://www.ffg.at/sites/default/files/allgemeine_downloads/thematische programme/Energie/EvAALuation2_D4.2_Handbuch_final.pdf

Johansson, T. (2011). Hail the impossible: P-values, evidence, and likelihood. Scandinavian Journal of Psychology, 52(2), 113–125. https://doi.org/10.1111/j.1467-9450.2010.00852.x

Kassambara, A. (2019). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. Abgerufen von: https://CRAN.R-project.org/package=rstatix

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. Journal of Statistical Software, 82(13).

Laugwitz, B.; Schrepp, M. & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In A. M. Heinecke & H. Paul (Hrsg.), Mensch & Computer 2006 - Mensch und Computer im Strukturwandel. München: Oldenbourg Verlag.

Little, R. J. (1992). Regression with missing X's: a review. Journal of the American Statistical Association, 87(420), 1227-1237.

Neubauer, A. B., & Voss, A. (2016). Validation and Revision of a German Version of the Balanced Measure of Psychological Needs Scale. Journal of Individual Differences, 37, 56–72. doi:10.1027/1614-0001/a000188

Morris, M. E., Miller, K., Ozanne, E., Hansen, R., Pearce, A. J., Santamaria, N., … Adair, B. (2013). Smart-Home Technologies to Assist Older People to Live Well at Home. Journal of Aging Science, 1. doi:10.4172/2329-8847.1000101

Müller, K. (2017). Here: A Simpler Way to Find Your Files. Abgerufen von: https://CRAN.R-project.org/package=here

Ooms, J. (2014). The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. arXiv:1403.2805. Abgerufen von: https://arxiv.org/abs/1403.2805

Pedersen, T. L. (2019). ggforce: Accelerating „ggplot2“. Abgerufen von: https://CRAN.R-project.org/package=ggforce

Peek, Aarts, S., & Wouters, E. J. M. (2015). Can Smart Home Technology Deliver on the Promise of Independent Living? A Critical Reflection Based on the Perspectives of Older Adults. Can Smart Home Technology Deliver on the Promise of Independent Living. In J. v. Hoof; G. Demiris; Eveline J. M. Wouters (Hrsg.), Handbook of Smart Homes, Health Care and Well-Being. Cham: Springer. doi:10.1007/978-3-319-01904-8_41-1

Reeder, B., Meyer, E., Lazar, A., Chaudhuri, S., Thompson, H. J., & Demiris, G. (2013). Framing the evidence for health smart homes and home-based consumer health technologies as a public health intervention for independent aging: A systematic review. International Journal of Medical Informatics, 82 (7), 565–579. doi:10.1016/j.ijmedinf.2013.03.007

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Revelle, W. (2019). psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University. Abgerufen von: https://CRAN.R-project.org/package=psych

Robinson, D., & Hayes, A. (2019). broom: Convert Statistical Analysis Objects into Tidy Tibbles. Abgerufen von: https://CRAN.R-project.org/package=broom

Schrepp, M., Thomaschewski, J., & Hinderks, A., (2018). UEQ User Experience Questionnaire. Abgerufen von: https://www.ueq-online.org/

Silge, J., & Robinson, D. (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. JOSS, 1(3). Abgerufen von: https://doi.org/10.21105/joss.00037

Slowikowski, K. (2019). ggrepel: Automatically Position Non-Overlapping Text Labels with „ggplot2“. Abgerufen von: https://CRAN.R-project.org/package=ggrepel

Statistik Austria, 2019. Bevölkerung nach Alter und Geschlecht. Abgerufen von: https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/bevoelkerungsstruktur/bevoelkerung_nach_alter_geschlecht/index.html

Teetor, N. (2018). zeallot: Multiple, Unpacking, and Destructuring Assignment. Abgerufen von: https://CRAN.R-project.org/package=zeallot

Tennekes, M. (2019). tmaptools: Thematic Map Tools. Abgerufen von: https://CRAN.R-project.org/package=tmaptools

Torchiano, M. (2019). effsize: Efficient Effect Size Computation. Abgerufen von: https://doi.org/10.5281/zenodo.1480624

Tierney, N. J., Harden, F. A., Harden, M. J., & Mengersen, K. L. (2015). Using decision trees to understand structure in missing data. BMJ Open, 5(6).

Warnes, G.R., Bolker, B., Lumley, T., & Johnson, R.C. (2018). gmodels: Various R Programming Tools for Model Fitting. Abgerufen von: https://CRAN.R-project.org/package=gmodels

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer: New York. Abgerufen von: https://ggplot2.tidyverse.org

Wickham, H. (2019). forcats: Tools for Working with Categorical Variables (Factors). Abgerufen von: https://CRAN.R-project.org/package=forcats

Wickham, H., & Bryan, J. (2019). readxl: Read Excel Files. Abgerufen von: https://CRAN.R-project.org/package=readxl

Wickham, H., François, R., Henry, L., & Müller, K. (2019). dplyr: A Grammar of Data Manipulation. Abgerufen von: https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2019). tidyr: Tidy Messy Data. Abgerufen von: https://CRAN.R-project.org/package=tidyr

Wickham, H., Hester, J., & François, R. (2018). readr: Read Rectangular Text Data. Abgerufen von: https://CRAN.R-project.org/package=readr

Wickham, H., & Seidel, D. (2019). scales: Scale Functions for Visualization. Abgerufen von: https://CRAN.R-project.org/package=scales