

<b>Konsortialführung:</b>	Albert-Ludwigs-Universität Freiburg
<b>Förderkennzeichen:</b>	01NVF18030
<b>Akronym:</b>	GET Sleep
<b>Projekttitel:</b>	GET Sleep - Stepped Care Modell für die Behandlung von Schlafstörungen
<b>Autoren:</b>	Prof. (apl.) Dr. Dr. Kai Spiegelhalder (Universitätsklinikum Freiburg) Prof. Dr. Harald Baumeister (Universität Ulm) Prof. Dr. Morten Moshagen (Universität Ulm)
<b>Förderzeitraum:</b>	1. Oktober 2019 – 30. September 2023

## Inhaltsverzeichnis

I. Abkürzungsverzeichnis.....	2
II. Abbildungsverzeichnis .....	2
III. Tabellenverzeichnis .....	2
1. Ziele der Evaluation .....	3
2. Darstellung des Evaluationsdesigns .....	6
3. Ergebnisse der Evaluation .....	14
4. Schlussfolgerungen und Empfehlungen des Evaluators .....	20
5. Literaturverzeichnis .....	20
6. Anhang.....	24
7. Anlagen.....	24

## I. Abkürzungsverzeichnis

ARR	Absolute Risk Reduction
AQoL-8D	Assessment of Quality of Life 8-Dimension
CSQ-I	Client Satisfaction Questionnaire (für internetbasierte Interventionen adaptiert)
GAD-7	Generalized Anxiety Disorder Scale-7
GKV	Gesetzliche Krankenversicherung
ICC	Intraklassenkorrelation
IG	Interventionsgruppe
ISI	Insomnia Severity Index
KVT-I	Kognitive Verhaltenstherapie für Insomnie
NEQ	Negative Effect Questionnaire
NNT	Number-needed-to-treat
PSQI	Pittsburgh Sleep Schlafqualitätsindex
QALY	Qualitätsbereinigter Lebensjahre
QIDS-SR16	Quick Inventory of Depressive Symptomatology
RCI	Reliable Change Index (RCI)
Skip	Die erste Stufe des Versorgungsmodells (hausärztliche Behandlung) wird übersprungen
SSS-8	Somatic Symptom Scale-8
T0	Erhebung zur Baseline
T1	Erhebung 4 Wochen nach Studieneinschluss
T2	Erhebung 12 Wochen nach Studieneinschluss
T3	Erhebung 6 Monate nach Studieneinschluss
TAI-OT	Technological Alliance Inventory für internetbasierte Intervention
TAU	Treatment as Usual
TiC-P	Trimbos/iMTA questionnaire for costs associated with psychiatric illness
WAI-I	Working Alliance Inventory for guided Internet interventions

## II. Abbildungsverzeichnis

Abbildung 1      Graphische Darstellung des Stepped Care Modells      S. 5

## III. Tabellenverzeichnis

Nicht vorhanden.

## 1. Ziele der Evaluation

Die Insomnie ist durch Ein-/Durchschlafstörungen und/oder Früherwachen gekennzeichnet, was zu erheblichen Beeinträchtigungen während des Tages führt (Morin et al., 2015) und mit einer verminderten Lebensqualität einhergeht (Kyle et al., 2010). Zudem gilt die Insomnie als Risikofaktor für die Entwicklung anderer psychischer Störungen, insbesondere Depressionen und Angststörungen (Hertenstein et al., 2019) sowie somatischer Erkrankungen wie Herz-Kreislauf-Erkrankungen (Li et al., 2014). Auch aus gesundheitsökonomischer Perspektive stellt die Insomnie eine Herausforderung dar, da durch Arbeitsausfälle und verringerte Produktivität die geschätzten jährlichen Kosten in Deutschland bei etwa 5900 Euro pro betroffener Person liegen (Thiart et al., 2016). Bei einer angenommenen Prävalenz von 5% (Schlack et al., 2013), belaufen sich die geschätzten jährlichen indirekten Kosten der Insomnie somit in Deutschland auf 25 Milliarden Euro, was die Relevanz dieser Störung unterstreicht.

Nationale und internationale Leitlinien empfehlen die kognitive Verhaltenstherapie für Insomnie (KVT-I) als Behandlung der ersten Wahl (Qaseem et al., 2016; Riemann et al., 2017; Riemann et al., 2023). Allerdings haben nur wenige Patient:innen Zugang zu dieser Behandlung. Stattdessen werden viele Patient:innen langfristig mit Benzodiazepin-rezeptoragonisten oder sedierenden Antidepressiva behandelt (Buth et al., 2019), was potenziell schädlich sein könnte und darüber hinaus nicht den Leitlinienempfehlungen entspricht (Qaseem et al., 2016; Riemann et al., 2017; Riemann et al., 2023).

Die Verbreitung von KVT-I ist eine bedeutende Herausforderung im Gesundheitswesen. Internetbasierte KVT-I könnte dabei helfen, die Versorgung von Insomnie zu verbessern. Meta-analytische Evidenz deutet darauf hin, dass internetbasierte KVT-I eine wirksame Alternative für die Bereitstellung von konventioneller KVT-I sein könnte (Gao et al., 2022; Hasan et al., 2022; van Straten et al., 2018). Allerdings ist unklar, wie internetbasierte KVT-I-Programme effektiv in bestehende Gesundheitssysteme integriert werden können, in denen Hausärztinnen und Hausärzte die Koordination der Patientinnen- und Patientenversorgung übernehmen. Einen möglichen Ansatz hierfür könnten gestufte Versorgungsmodelle bieten. In Anbetracht der zentralen Rolle der Hausärztinnen und Hausärzte in Deutschland könnte eine hausärztliche Erstbehandlung und Diagnostik die erste Stufe in einem solchen Versorgungsmodell darstellen, bei der bei Bedarf eine internetbasierte KVT-I als zweite Stufe und falls notwendig in der dritten Stufe eine fachärztliche Behandlung erfolgt. Aktuell gibt es erst wenige Untersuchungen zu gestuften Versorgungsmodellen für die Behandlung der Insomnie (Forsell et al., 2019; Salomonsson et al., 2018; Zhou et al., 2020) und keine dieser Studien beinhaltet eine Behandlung durch Hausärztinnen und Hausärzte. Darüber hinaus stellt sich bezüglich der internetbasierten KVT-I die Frage, was das optimale Ausmaß der therapeutischen Begleitung ist.

Im Rahmen von GET Sleep wurde ein sektorenübergreifendes Stepped Care Modell für die Behandlung von Insomnien implementiert. Dabei wurden Hausarztpraxen geschult, die Behandlung in Übereinstimmung mit den entsprechenden Leitlinien in einem gestuften Behandlungspfad mit den folgenden Stufen zu organisieren: 1. Hausärztliche standardisierte Diagnostik und psychoedukative Erstbehandlung; 2. Internet- und mobil-basierte Kognitive Verhaltenstherapie für Insomnie (KVT-I); 3. Psychotherapeutische, psychiatrische oder schlafmedizinische Fachbehandlung. Die bei der Schulung der Hausarztpraxen verwendeten Materialien sind in Anlage 1 des Ergebnisberichts zusammengestellt. Folgend eine detaillierte Beschreibung der drei Stufen des Versorgungsmodells (siehe auch Abbildung 1).

### 1) Hausärztliche standardisierte Diagnostik und psychoedukative Erstbehandlung

Das Screening von Patientinnen und Patienten erfolgte durch die Hausärztinnen und Hausärzte mittels kurzem Fragebogen, der den Hausarztpraxen durch die Konsortialführung zur Verfügung gestellt wurde (siehe Anlage 1 des Ergebnisberichts). Auf Basis der Ergebnisse und, bei Bedarf, einer Blutentnahme (kleines Blutbild: 1,10 €, TSH: 3,70 €,  $\gamma$ -GT: 0,25 €) haben die Hausärztinnen und Hausärzte die Patientinnen und Patienten in das Projekt

eingeschlossen. Für diese Rekrutierung von Studienteilnehmenden durch die Hausärztinnen und Hausärzte war ein Zeitbedarf von 60 min vorgesehen, der nach in Anlehnung an EBM-Ziffer 35210113 mit 51,60 € vergütet wurde. Die daran anschließende hausärztliche psychoedukative Erstbehandlung beinhaltete die Vermittlung der so genannten „schlafhygienischen Regeln“ sowie basale Informationen zu den Themen Schlaf und Schlafstörungen, wofür ein Arbeitsaufwand von 30 min pro Patientin/ Patient vorgesehen war. Die Vergütung hierfür sowie für maximal einen weiteren 30-minütigen Termin der Hausärztinnen und Hausärzte mit den Patientinnen/Patienten im Rahmen des Stepped Care Modells wurde in Anlehnung an EBM-Ziffer 35210113 mit 2 x 25,80 € vergütet (=51,60 €). Der Behandlungserfolg dieser Intervention wurde nach zwei bis vier Wochen durch die Hausärztinnen und Hausärzte kontrolliert. Bei nach klinischem Eindruck fehlendem Behandlungserfolg wurden Patientinnen und Patienten der 2. Stufe des Modells zugewiesen. Aber auch die Patientinnen und Patienten konnten sich bei subjektiv ausbleibendem Behandlungserfolg direkt für die 2. Stufe des Stepped Care Modells anmelden. Vor Durchführung der psychoedukativen Erstbehandlung konnten die Hausärztinnen und Hausärzte auch entscheiden, dass die 1. Stufe des Versorgungsmodells übersprungen wird. Diese direkte Zuweisung zur 2. Stufe des Versorgungsmodells sollte insbesondere erfolgen, falls kein substanzieller Effekt der Behandlung in der 1. Stufe auf die schlafbezogene Symptomatik zu erwarten war.

## 2) Internet- und mobil-basierte KVT-I

Die eingesetzte therapeutengestützte internet- und mobil-basierte Behandlung GET.ON Insomnie des GET.ON Instituts (siehe Ergebnisbericht Anlage 2 für eine Darstellung der technischen Entwicklungen) beruht auf der KVT-I und umfasst somit Psychoedukation, Entspannungstechniken, Schlafrestriktion, Stimuluskontrolle und kognitive Interventionen (siehe Spiegelhalder et al., 2011). Darüber hinaus enthält die Intervention eine Einheit zur Rückfallprävention sowie optional wählbare Module zur Bewältigung komorbider Erkrankungen. Die Intervention ist in acht Einheiten aufgeteilt, deren Bearbeitung jeweils ca. 45-60 Minuten dauert. Die Bearbeitung von einem Modul pro Woche war vorgesehen. Um jedoch interindividuelle Unterschiede im therapeutischen Prozess zu berücksichtigen, hatten Patientinnen und Patienten die Möglichkeit, die Intervention langsamer oder schneller zu bearbeiten. Die Behandlung wurde zur Bestimmung einer möglichst optimalen Kosten-Effektivität in drei verschiedenen Intensitäts-Varianten durchgeführt, die sich im Ausmaß und in der Art der Unterstützung durch E-Coaches unterschieden. Alle Patientinnen und Patienten erhielten ein telefonisches/video-basiertes Erstgespräch zur Bedürfnisklärung sowie zum Beziehungs- und Motivationsaufbau sowie ein Abschlussgespräch, in dem gegebenenfalls Empfehlungen für eine weitere Behandlung gegeben wurden.

Die Variante GET.ON Insomnie Standard enthielt darüber hinaus ein intensives schriftliches Coaching durch E-Coaches mit einem Zeitaufwand von 8 x 25 min (Vergütung inklusive Erstgespräch und Abschlussgespräch: 490,72 €). Dabei gaben die E-Coaches zu jeder der acht Einheiten der Intervention ein schriftliches Feedback. Die E-Coaches haben sich hierfür alle Angaben, die die Teilnehmenden im Rahmen der Einheit machten, durchgelesen und hierzu konkretes, personalisiertes Feedback bezogen auf die jeweilige Situation der Teilnehmenden gegeben. Die Variante GET.ON Insomnie Flex zeichnete sich durch ein adhärenz-fokussiertes schriftliches Coaching durch E-Coaches mit einem mittleren geschätzten Zeitaufwand von insgesamt 50 min aus (Vergütung inklusive Erstgespräch und Abschlussgespräch: 221,92 €). Dabei erfolgte dieses adhärenz-fokussierte schriftliche Coaching jeweils auf Anfrage der Patientinnen und Patienten. Derartige Anfragen konnten per Nachrichtenfunktion gestellt werden, wobei das Feedback der E-Coaches dann individualisiert mit Fokus auf die spezifischen vorgebrachten Fragen oder Probleme gegeben wurde. In der Variante GET.ON Insomnie Basic war keine weitere Unterstützung durch E-Coaches vorgesehen (Vergütung inklusive Erstgespräch und Abschlussgespräch: 132,32 €). Die Handbücher für die E-Coaches sind in Anlage 3 des Ergebnisberichts dargestellt.

Die E-Coaches waren im Rahmen eines Anstellungsverhältnisses beim Konsortialpartner GET.ON Institut für Online Gesundheitstrainings GmbH beschäftigt. Es waren 11 E-Coaches beteiligt, welche einen Diplom- oder Masterabschluss in Psychologie und darüber hinaus eine zusätzliche absolvierte bzw. weit fortgeschrittene Ausbildung zum Psychologischen Psychotherapeuten aufwiesen. Die Kontaktaufnahme zu den Projektteilnehmenden fand per Telefon (Eingangs- und Abschlussgespräch) sowie über die verschlüsselte Plattform (im Rahmen der Nachrichtenfunktion) statt.

Der Behandlungserfolg der digitalen Intervention wurde den Hausärztinnen und Hausärzten in einem Befundbericht übermittelt (Vergütung: 8,34 €), so dass im Bedarfsfall die Zuweisung zu einer psychotherapeutischen, psychiatrischen oder schlafmedizinischen Fachbehandlung (3. Stufe des Stepped Care Modells) erfolgen konnte. Die behandelnden Hausärztinnen und Hausärzte entschieden über das weitere Vorgehen. Für diese Entscheidung erhielten die Hausarztpraxen einen Bericht vom E-Coach der internetbasierten Intervention. Dieser umfasste den Post-Treatment Insomnia Severity Index-Gesamtscore (ISI; Bastien et al., 2001), der außerhalb des Forschungsprozesses auf der Behandlungsplattform erhoben wurde sowie eine Empfehlung darüber, ob und von wem die Behandlung nach der 2. Stufe fortgesetzt werden sollte. Die Hausärztinnen und Hausärzte erhielten dabei die generelle Empfehlung, Patientinnen und Patienten mit einem ISI-Gesamtscore  $\geq 15$  und einer komorbiden psychischen Erkrankung an eine psychiatrische oder psychotherapeutische Fachbehandlung zu überweisen. Alle anderen Patientinnen und Patienten mit einem ISI-Gesamtscore  $\geq 15$  sollten an eine schlafmedizinische Fachbehandlung überwiesen werden. Alle Daten wurden auf ISO 27000-zertifizierten Servern sicher gespeichert und eine sichere Datenübertragung wurde durch Hypertext Transfer Protocol Secure (HTTPS) mit Secure Sockets Layer (SSL) Zertifikaten (AES-256 und SHA-1, 2048-bit RSA) gewährleistet.

### 3) Psychotherapeutische, psychiatrische oder schlafmedizinische Fachbehandlung

Die psychotherapeutische, psychiatrische oder schlafmedizinische Fachbehandlung erfolgte im Rahmen der Regelversorgung durch entsprechende Fachkräfte, die auf die Behandlung von Insomnien spezialisiert sind.



**Abbildung 1:** Graphische Darstellung des Stepped Care Modells.

**Zielgruppe:** Die Zielgruppe der neuen Versorgungsform waren erwachsene Patientinnen und Patienten mit der Diagnose nicht-organische Insomnie (ICD-10: F51.0) oder Ein- und Durchschlafstörungen (ICD-10: G47.0).

**Interoperabilität und offenen Schnittstellen:** Der Studien-Code bildete das Kernstück der studieninternen Interoperabilität zwischen den Konsortialpartnern. Die Interventionsplattform der GET.ON Instituts folgt gängigen Prinzipien der Interoperabilität und ermöglicht einen Datenexport basierend auf dem von HL7 veröffentlichten FHIR-Standard. Sie entspricht der veröffentlichten FHIR® Spezifikation v4.0.1: R4 ohne zusätzliche Profilierung.

## 2. Darstellung des Evaluationsdesigns

Die drei Varianten der sektorenübergreifenden neuen Versorgungsform, die sich in der Intensität der therapeutischen Begleitung während der internetbasierten KVT-I unterschieden, wurden im Vergleich mit der Treatment-as-Usual-Kontrollgruppe (TAU) in einer vier-armigen cluster-randomisierten kontrollierten Studie untersucht. Zur Vermeidung von Cross-over-Effekten wurden Hausarztpraxen als Einheit der Randomisierung gewählt. Aus ethischen Gründen und zur Kostenreduktion wurden die Cluster im Verhältnis 3:3:3:1 (IG1: IG2: IG3: TAU) randomisiert.

Im veröffentlichten Studienprotokoll und im ursprünglich eingereichten Evaluationskonzept war die Studienteilnahme von vier telemedizinisch arbeitenden Hausarztpraxen geplant, welche jedoch aufgrund von Rekrutierungsschwierigkeiten von Vor-Ort-Hausarztpraxen auf 17 erhöht wurde. Des Weiteren war eine Unterscheidung der Stichprobe in vier Gruppen (i.e., Insomnie ohne psychische und körperliche Komorbidität, Insomnie mit psychischer Komorbidität, Insomnie mit körperlicher Komorbidität, Insomnie mit psychischer und körperlicher Komorbidität) vorgesehen. Aufgrund der Rekrutierungsschwierigkeiten wurde entschieden, diese Unterscheidung aufzugeben (siehe Ausführliches Evaluationskonzept GET Sleep; Version 3.0, 16.12.2022).

### 2.1. Ein- und Ausschlusskriterien

Teilnehmende Hausärztinnen und Hausärzte mussten eine Kassenzulassung bzw. eine Abrechnungsberechtigung mit den gesetzlichen Krankenkassen haben. Teilnehmende Vor-Ort-Hausärztinnen und Hausärzte mussten ihren Sitz in Bayern oder Baden-Württemberg haben und konnten Patientinnen und Patienten aus diesen Bundesländern rekrutieren. Teilnehmende telemedizinisch arbeitende Hausarztpraxen konnten Patientinnen und Patienten aus ganz Deutschland rekrutieren. Die teilnehmenden Hausärztinnen und Hausärzten sind dem Selektivvertrag beigetreten, der im Projekt geschlossen wurde.

Teilnehmende Patientinnen und Patienten mussten folgende Ein- und Ausschlusskriterien erfüllen, die im Rahmen eines hausärztlichen Gesprächs überprüft wurden, wobei den Hausärztinnen und Hausärzten hierfür zur optionalen Nutzung ein Screening-Fragebogen zur Verfügung gestellt wurde (siehe Anlage 1 des Ergebnisberichts):

Einschlusskriterien: Diagnose einer nicht-organischen Insomnie (ICD-10: F51.0) oder von Ein- und Durchschlafstörungen (ICD-10: G47.0).

Ausschlusskriterien: Unbehandeltes Schlafapnoe-Syndrom (ICD-10: G47.3); unbehandeltes Restless Legs Syndrom/Syndrom der nächtlichen periodischen Beinbewegungen (ICD-10: G25.8); aktuell laufende Psychotherapie für Insomnie; Erkrankungen, die durch die KVT-I verschlechtert werden können (bipolare affektive Störung, ICD-10: F31.x; Epilepsie, ICD-10: G40.x); Erkrankungen, die die zuverlässige Nutzung der internet- und mobil-basierten Behandlung gefährden (z. B. organische, einschließlich symptomatischer psychischer Störungen, ICD-10: F00-F09; Psychische und Verhaltensstörungen durch psychotrope Substanzen, ICD-10: F10-F19; Schizophrenie, schizotype und wahnhaftige Störungen, ICD-10: F20-F29).

## 2.2. Fallzahlbegründung

Da angenommen wurde, dass sich die TAU-Gruppe in den Endpunkten deutlich stärker von den IGs unterscheidet als die IGs untereinander, wurden die Patientinnen und Patienten aus ethischen Gründen und zur Kostenreduktion im Verhältnis 3:3:3:1 (IG1:IG2:IG3:KG) randomisiert. Zunächst wurde dabei eine Stichprobengröße von 4.268 Patientinnen und Patienten angestrebt, da die differentielle Wirksamkeit der neuen Versorgungsform innerhalb von vier Gruppen von Patientinnen und Patienten, die sich im Vorliegen von Komorbiditäten unterscheiden, untersucht werden sollte. Die damalige Fallzahlbegründung beruhte auf folgenden Annahmen: 320 teilnehmende Hausarztpraxen mit einem Median der Rekrutierungsleistung von 9 Patientinnen und Patienten bei einer Standardabweichung von 14 Patientinnen und Patienten, einem ICC von 0,02, einer Korrelation der T3-Werte mit den T0-Werten von  $r=0,5$ , einem  $\alpha$  von 0,05 und einer Teststärke ( $1-\beta$ ) von 80%. Daraus ergab sich eine notwendige Stichprobengröße von 1.067 (IGs jeweils  $n=320$ , KG  $n=107$ ) für jede Gruppe aus Patientinnen und Patienten bei einer angenommenen Effektstärke von  $d = 0,25$ . Im Verlaufe des Projekts wurde deutlich, dass diese Stichprobengröße realistischerweise nicht erreicht werden konnte, weshalb, wie im Dokument Ausführliches Evaluationskonzept GET Sleep (Version 3.0., 16.12.2022) beschrieben, die geplante Unterscheidung der vier Patientengruppen auf Basis vorliegender Komorbiditäten aufgegeben wurde. Dies reduzierte die erforderliche Stichprobengröße auf insgesamt 1.067 Patienten. Aufgrund des Projektendes wurde die Rekrutierung am 30. September 2022 beendet, obwohl die angestrebte Fallzahl bis zu diesem Zeitpunkt noch nicht erreicht worden war.

## 2.3. Randomisierung

Die Randomisierung erfolgte auf der Ebene der Hausärztinnen und Hausärzte. Es erfolgte eine balanciert permutierte Blockrandomisierung (neun Strata auf Grundlage der Bevölkerungsdichte und des Durchschnittseinkommens in den Einzugsgebieten der jeweiligen Hausarztpraxen, ein Stratum für telemedizinisch arbeitende Hausarztpraxen) mit unterschiedlichen Blocklängen, die der Konsortialführung verdeckt blieben, um einen Selektionsbias zu verhindern. Hausärztinnen und Hausärzte aus Gemeinschaftspraxen wurden jeweils in den selben Studienarm randomisiert. Die Randomisierung erfolgte durch Dr. Martina Bader und Prof. Dr. Morten Moshagen, die ansonsten bis zu der Datenauswertung nicht an der Studie beteiligt waren und daher für alle Vorgänge in der Studie verblindet waren. Patientinnen und Patienten vereinbarten die Termine bei den jeweiligen Hausarztpraxen ohne zuvor Informationen über deren Gruppenzugehörigkeit zu haben. Die Hausärztinnen und Hausärzte wurden angewiesen, ihre Patientinnen und Patienten die Gruppenzuordnung nicht mitzuteilen, bis diese die Eingangsbefragung (T0) abgeschlossen hatten. Nach Abschluss der Eingangsbefragung wurden die Patientinnen und Patienten auf die Interventionsplattform weitergeleitet und dort über ihre Studiengruppe informiert. Zusätzlich erhielten die Patientinnen und Patienten eine E-Mail mit Information über die Gruppenzugehörigkeit. Ein Informed Consent zur Studienteilnahme war zwingend erforderlich. Weder Patientinnen/Patienten noch Behandelnde (Hausärztinnen/Hausärzte, E-Coaches) waren verblindet.

## 2.4. Datenerhebung

Die Erhebung aller psychometrischen und gesundheitsökonomischen Outcome-Parameter erfolgte online durch die Befragungssoftware LimeSurvey (<https://www.limesurvey.org/>) zu den Messzeitpunkten T0 (Baseline), T1 (nach 4 Wochen), T2 (nach 12 Wochen) und T3 (nach 6 Monaten). Der primäre Outcome-Parameter war die Schwere der Insomnie zum Messzeitpunkt T3, erfasst mit dem Insomnia Severity Index (ISI, Bastien et al., 2001). Der ISI besteht aus sieben Items auf einer 5-stufigen Likert-Skala (0-4 Punkte; Summenscore 0-28 Punkte), die den wahrgenommenen Schweregrad der Insomnie während der vorangegangenen zwei Wochen abbilden. Mehrere Studien belegen die psychometrische

Güte des ISI, Cronbachs Alpha lag in diesen Studien zwischen 0,70 und 0,90 (Bastien et al., 2001; Gagnon et al., 2013; Morin et al., 2011).

Folgende sekundäre Outcome-Parameter wurden erhoben: Zur Erfassung der Schlafqualität wurde der Pittsburgh Schlafqualitätsindex (PSQI; Buysse et al., 1989) verwendet. Der PSQI besteht aus 19 Items und erfasst verschiedene Aspekte der Schlafqualität. Der Summenscore liegt zwischen 0 und 21 und die interne Konsistenz bei 0,8 (Carpenter & Andrykowski, 1998). Die Beurteilung der Lebensqualität erfolgte mit dem Assessment of Quality of Life 8-Dimension (AQoL-8D; Richardson et al., 2014). Unter Verwendung der Trade-Off Methode kann der AQoL-8D zur Schätzung qualitätsbereinigter Lebensjahre (QALYs) herangezogen werden. Der AQoL-8D weist mit einem Cronbachs Alpha von 0,96 eine ausgezeichnete interne Konsistenz auf (Richardson et al., 2014). Depressive Symptome wurden mit dem Quick Inventory of Depressive Symptomatology, self-report erhoben (QIDS-SR16; Rush et al., 2003). Der QIDS-SR16 weist eine gute interne Konsistenz auf (Cronbachs Alpha=0,86; Trujols et al., 2018). Das Vorliegen einer Depression zum Messzeitpunkt T3 wurde definiert als ein Score  $\geq 13$  auf dem QIDS-SR16 (Lamoureux et al., 2010). Angstsymptome wurden mit der Generalized Anxiety Disorder Scale-7 erfasst (GAD-7; Löwe et al., 2008), die eine gute interne Konsistenz aufweist (Cronbachs Alpha=0,89; Spitzer et al., 2006). Somatische Symptome wurden durch die Somatic Symptom Scale-8 erfasst (SSS-8; Gierk et al., 2014; Cronbachs Alpha=0,81). Für die gesundheitsökonomische Evaluation wurde der retrospektive Selbstbericht-Fragebogen Trimbo/iMTA questionnaire for costs associated with psychiatric illness (TiC-P) (Bouwman et al., 2013; Buntrock et al., 2021; Hakkaart-van Roijen et al., 2002) verwendet, dessen Test-Retest-Reliabilität zufriedenstellend ist (Bouwman et al., 2013).

Als interventionsbezogene Variablen wurde in der T2-Erhebung bei allen Patientinnen und Patienten, die an der 2. Stufe des Versorgungsmodells teilgenommen haben, das Working Alliance Inventory for guided Internet interventions erhoben (WAI-I; Gómez Penedo et al., 2020; Cronbachs Alpha=0,93), um die Qualität der Allianz zwischen Patientinnen/Patienten und E-Coach zu erfassen. Um die technologische Allianz zwischen Patientinnen/Patienten und der internetbasierten Intervention zu erfassen wurde weiterhin das Technological Alliance Inventory verwendet (TAI-OT; Herrero et al., 2020). Bei allen Patientinnen und Patienten wurde darüber hinaus der für internetbasierte Interventionen adaptierte Client Satisfaction Questionnaire erhoben (CSQ-I; Attkisson & Zwick, 1982; Boß et al., 2016; Cronbachs Alpha=0,93) erhoben. Um mögliche negative Effekte der Intervention zu erfassen wurde der Negative Effect Questionnaire (NEQ; Rozental et al., 2019; Cronbachs Alpha=0,95) sowie ein eigens im Projekt entwickelter Fragebogen für die Nebenwirkungen der KVT-I (NEQ-Insomnia; Anlage 5 des Ergebnisberichts) verwendet.

Darüber hinaus war es bei der Konzeption der Studie beabsichtigt, mögliche Moderations- und Mediationseffekte zu untersuchen. Neben den soziodemographischen Variablen wurden hierfür zu den Messzeitpunkten T0, T1, T2 und T3 die folgenden Skalen erhoben: Dysfunctional Beliefs and Attitudes about Sleep Scale (DBAS-10; Cronbach's Alpha=0,69; Morin et al., 1993; Espie et al., 2000), die Pre-Sleep Arousal Scale (PSAS; Cronbach's Alpha=0,80-0,94; Nicassio et al., 1985; Giesermann et al., 2012), das Brief Fatigue Inventory (Cronbach's Alpha=0,96; Mendoza et al., 1999), die Perceived Stress Scale (PSS; Cronbach's Alpha=0,78; Cohen et al., 1983), der Sleep Hygiene Index (SHI; Cronbach's Alpha=0,66; Mastin et al., 2006) und der Cognitive Emotion Regulation Questionnaire (CERQ-short; Cronbach's Alpha=0,68-0,81; Garnefski & Kraaij, 2006).

Fehlende Werte auf den Skalen-Gesamtscores wurden mittels Multipler Imputation (predictive mean matching) mit dem R-Paket mice (van Buuren & Groothuis-Oudshoorn, 2011) imputiert. Es wurden Imputationsmodelle ohne Mehrebenenstruktur verwendet, da Mehrebenen-Imputationsmodelle zur Berücksichtigung der Cluster-Struktur keine Konvergenz erreichten, was vermutlich auf die geringe Anzahl an Clustern zurückzuführen war. Die Cluster-Zugehörigkeit wurde dennoch berücksichtigt, indem diese als Prädiktor in



allen Imputationsmodellen aufgenommen wurde. Ebenso waren Geschlecht und Alter der Patientinnen und Patienten sowie die Gesamtscores auf der jeweiligen Skala zu allen früheren Messzeitpunkten Prädiktoren in allen Imputationsmodellen. Weitere Prädiktoren wurden datengetrieben ausgewählt, wobei die minimale Korrelation zwischen Prädiktoren und dem Kriterium des Imputationsmodells auf .30 und der minimale Anteil an nutzbaren Fällen auf .25 festgelegt wurde. Es wurden 10 Datensätze imputiert.

Barrieren und Förderfaktoren der neuen Versorgungsform wurden anhand einer qualitativen Inhaltsanalyse untersucht. Hierfür wurden in den drei Stakeholdergruppen (Patientinnen/Patienten, Hausärztinnen/Hausärzte und E-Coaches) Interviews geführt. Anhand von folgenden Charakteristika wurde zunächst in der Gruppe der Patientinnen und Patienten ein theoretisches Sampling durchgeführt: Responder (hier definiert als eine Veränderung von  $\geq 8$  Punkten im ISI-Gesamtscore von T0 zu T3 (Morin et al., 2011), Dropouts (d.h., Personen, die weniger als 80% der internetbasierten KVT-I innerhalb von 12 Wochen bearbeitet haben), Alter und Studiengruppe.

Für die Patientinnen und Patienten wurde vorab keine Stichprobengröße für die Interviews definiert; stattdessen wurde nach dem Prinzip der theoretischen Sättigung rekrutiert (Patton, 2002; Sandelowski, 1995). Dementsprechend wurde die Datenerhebung so lange fortgesetzt, bis absehbar war, dass keine weiteren Erkenntnisse aus zusätzlichen Interviews gewonnen werden würden. Nach 22 Interviews mit Patientinnen und Patienten wurden die Interviews vorläufig begutachtet, um zu entscheiden, welche Charakteristika bei der folgenden Rekrutierung fokussiert werden sollten. Da angenommen wurde, dass der Rücklauf bei den Hausärztinnen und Hausärzten sehr gering sein würde, wurden alle 46 teilnehmenden Hausärztinnen und Hausärzte zu den qualitativen Interviews eingeladen. Aufgrund der geringen Anzahl an E-Coaches, die im Projekt tätig waren (n=11), wurden ebenfalls alle E-Coaches zu den qualitativen Interviews eingeladen.

Vor den Interviews wurden für die drei Stakeholdergruppen separate Interviewleitfäden basierend auf dem Hierarchical Model of Service Quality (Dagger et al., 2007) erstellt (siehe Anlage 6 des Ergebnisberichts). Die Interviewleitfäden adressieren die vier Dimensionen des Modells (Interpersonelle Qualität, Technische Qualität, Umgebungsqualität, Administrative Qualität). Die Interviewleitfäden wurden angelehnt an Helfferich (2011) formuliert.

Alle Interviews wurden von geschulten Psychologinnen (Mindestqualifikation: abgeschlossenes Bachelorstudium) durchgeführt. Vor Durchführung der Interviews wurden anhand von Fallvignetten Probeinterviews geführt. Die angestrebte Dauer der Interviews betrug 45-60 Minuten für Patientinnen und Patienten, 30-45 Minuten für E-Coaches, und 30 Minuten für Hausärztinnen und Hausärzte.

Die Interviews wurden mit der Videokonferenzsoftware Zoom mit ausgeschalteter Kamerafunktion durchgeführt und mittels der Aufnahmefunktion von Zoom aufgezeichnet. Zur Transkription wurden einfache Transkriptionsregeln angewendet (d.h. das Audiomaterial wurde wörtlich transkribiert und im Wesentlichen so niedergeschrieben, wie gesprochen). Grammatikfehler und Füllwörter wurden beibehalten, jedoch wurden Dialektvariationen auf Standarddeutsch angepasst.

## 2.5. Effektevaluation

**Primäre Fragestellung:** Untersuchung der Wirksamkeit der neuen Versorgungsform im Vergleich zu TAU im Hinblick auf den Schweregrad der Insomnie.

**Arbeitshypothese:** Es wird erwartet, dass die drei Varianten der sektorenübergreifenden Versorgungsform zu einer signifikanten Verringerung der Insomnieschwere im Vergleich zu TAU führen. Bezüglich der Unterschiede zwischen den drei Varianten der sektorenübergreifenden Versorgungsform wurden keine spezifischen Annahmen formuliert.

**Statistische Analysen:** Die primäre Wirksamkeitsanalyse wurde auf der Grundlage aller Patientinnen und Patienten nach dem Intention-to-Treat-Prinzip durchgeführt. Die Effekte der Bedingung (IG1, IG2, IG3, TAU) auf das primäre Outcome Insomnieschwere (ISI zu T3) wurden mit linearen Regressionsmodellen untersucht. Alle Modelle enthielten das Alter und das Geschlecht der Patientinnen und Patienten sowie den Baseline-Wert der Insomnieschwere als Kovariaten. Abweichend vom Evaluationskonzept wurden keine Mehrebenen-Regressionsmodelle berechnet, da diese Modelle nicht konvergierten, was wahrscheinlich auf die geringe Anzahl an Clustern sowie dem geringen Anteil an Varianz, der durch die Clusterzugehörigkeit erklärt wurde, zurückzuführen ist. Die p-Werte wurden nach dem Bonferroni-Holm-Verfahren adjustiert.

Als zusätzliche Analysen für das primäre Outcome Insomnieschwere wurden analog zum Intention-to-Treat Prinzip Per-Protocol-Analysen durchgeführt basierend auf den Daten von Patientinnen und Patienten, die einen wesentlichen Teil der internetbasierten KVT-I (d.h., 80% der internetbasierten KVT-I) bearbeitet hatten. Darüber hinaus wurde für die Effektevaluation die reliable Reduktion des Schweregrads der Insomnie mit dem Reliable Change Index (RCI) nach Jacobson und Truax (Jacobson & Truax, 1992) berechnet. Basierend auf einer Validierungsstudie mit 410 Patientinnen und Patienten in der Primärversorgung (Gagnon et al., 2013) wurde für die Berechnung des RCIs ein Cronbach's Alpha von 0.92 angenommen. Folgend wurden Patientinnen und Patienten als Responder (d.h., reliable Veränderung im Schweregrad der Insomnie) und Non-Responder (d.h., keine reliable Veränderung des Schweregrads der Insomnie) kategorisiert. Anschließend wurde die Wahrscheinlichkeit für eine Response mittels logistischer Regression zwischen den Bedingungen verglichen. Ferner wurde die klinische Signifikanz anhand von number-needed-to-treat (NNT)-Analysen ermittelt (Laupacis et al., 1988). Hierfür wurde zunächst der ISI-Gesamtscore zu T3 in eine dichotome Variable überführt (Patientinnen und Patienten mit einem ISI Score  $\leq 7$  wurden als remittiert klassifiziert). Folgend wurde die NNT wird als Inverse der Absolute Risk Reduction (ARR) berechnet.

**Sekundäre Fragestellung:** Untersuchung der Wirksamkeit der drei Varianten der sektorenübergreifende Versorgungsform im Vergleich zu TAU bezogen auf die Schlafqualität, Lebensqualität, depressive Symptome, Angstsymptome, somatische Symptome sowie das Vorliegen einer depressiven Symptomatik nach der Behandlung.

**Arbeitshypothese:** Für diese Fragestellungen wurden keine spezifischen Arbeitshypothesen formuliert, da sie explorativ untersucht wurden (d.h. die Studie nicht auf die Beantwortung dieser Fragestellungen, sondern auf die primäre Fragestellung gepowert war).

**Statistische Analysen:** Analog zu den Intention-to-Treat und Per-Protocol-Analysen bei dem primären Outcome Insomnieschwere wurden lineare Regressionsmodelle für folgende Outcome-Parameter zu T3 spezifiziert: PSQI, AQoL-8D, QIDS-SR16, GAD-7 und SSS-8. Darüber hinaus wurde der Effekt der Studienbedingung auf das Vorliegen einer depressiven Symptomatik zu T3 mittels einer Poisson Regression mit robusten Standardfehlern untersucht (Zou, 2004).

**Moderations- und Mediationsanalysen:** Es war ursprünglich geplant, potentielle Moderatoren von Behandlungseffekten (Alter, Geschlecht, Baselinewerte in Bezug auf Depression, Angst, dysfunktionale Überzeugungen zum Thema Schlaf, Fatigue, Stress, Schlafhygiene und Emotionsregulationsstrategien) explorativ zu untersuchen. Darüber hinaus war es geplant, dysfunktionale Überzeugungen zum Thema Schlaf, Fatigue, Stress, Schlafhygiene und Emotionsregulationsstrategien sowie die therapeutische Beziehung und die technische Beziehung als potenzielle Mediatoren explorativ zu untersuchen.

**Arbeitshypothese:** Für diese Fragestellungen wurden keine spezifischen Arbeitshypothesen formuliert, da sie explorativ untersucht werden sollten (d.h. die Studie nicht auf die Beantwortung dieser Fragestellungen, sondern auf die primäre Fragestellung gepowert war).

Statistische Analysen: Die erreichte Stichprobengröße lag deutlich unter der erwarteten Stichprobengröße, so dass die statistische Power zum Aufdecken von Moderationseffekten sehr gering ist. Zudem zeigten sich in den primären Untersuchungen keine statistisch signifikanten Effekte der Interventionsgruppen auf die Schwere der Insomnie. Daher wurde von der Durchführung von Mediations- und Moderationsanalysen abgesehen. Stattdessen werden die Summenwerte dieser Variablen zu den jeweiligen Messzeitpunkten deskriptiv berichtet.

## 2.6. Prozessevaluation

**1. Fragestellung:** Welche Barrieren und Förderfaktoren bestehen für das neue Versorgungsmodell aus Sicht der Patientinnen und Patienten, E-Coaches und Hausärztinnen und Hausärzte?

Arbeitshypothese: Für diese Fragestellung wurde keine spezifische Arbeitshypothese formuliert, da die Fragestellung explorativ untersucht wurde.

Statistische Analyse: Die Barrieren und Förderfaktoren der drei Varianten der sektorübergreifenden Versorgungsform wurden mittels einer qualitativen Inhaltsanalyse untersucht (Kuckartz & Rädigker, 2022). Diese Methodik wurde gewählt, da sie eine Kombination aus deduktiver und induktiver Kategorienbildung ermöglicht. Zur Analyse der qualitativen Interviews wurde die Software MAXQDA (Version 20.8.0; VERBI Software, 2021) verwendet. Während der initialen Textarbeit wurden 27 Transkripte (51% aller Transkripte) durchgelesen. Daraufhin wurde ein Kategoriensystem für alle Stakeholder entwickelt, bei dem die Kategorien und Unterkategorien deduktiv basierend auf dem Hierarchical Model of Service Quality (Dagger et al., 2007) abgeleitet wurden. Falls erforderlich, wurden die Themenbereiche der Unterkategorien induktiv erweitert. Alle Interviews der E-Coaches und Hausärztinnen/Hausärzte sowie von 29 Patientinnen und Patienten wurden von zwei unabhängigen Personen codiert. Die verbleibenden 10 Interviews mit Patientinnen/Patienten wurden einfach codiert. Alle doppelt kodierte Interviews wurden verglichen, Abweichungen in den Kodierungen wurden durch gemeinsame Diskussion der Kodierenden geklärt. Textabschnitte, die nicht relevant für die Forschungsfrage waren, wurden nicht codiert. Textabschnitte, die mehrere Themenbereiche ansprachen, durften mehreren Kategorien zugeordnet werden. Anschließend wurden die kodierte Unterkategorien ausgewertet, um Barrieren und Förderfaktoren für die neue Versorgungsform zu identifizieren. Es wurden nur Aussagen ausgewertet, sofern sich mindestens ein Stakeholder zu dieser Unterkategorie geäußert hat.

**2. Fragestellung:** Wie ist die patientenseitige Zufriedenheit mit der Behandlung?

Arbeitshypothese: Für diese Fragestellung wurde keine spezifische Arbeitshypothese formuliert, da die Fragestellung explorativ untersucht wurde (d. h. die Studie nicht auf die Beantwortung dieser Fragestellungen, sondern auf die primäre Fragestellung gepowert war).

Statistische Analyse: Analog zu den Intention-to-Treat Analysen bei dem primären Outcome wurden für die patientenseitige Zufriedenheit ein lineares Regressionsmodell für den Outcome Parameter CSQ-I spezifiziert.

**3. Fragestellung:** Wie ist die therapeutische Allianz zu den E-Coaches und die technologische Allianz zur internetbasierten KVT-I?

Arbeitshypothese: Für diese Fragestellung wurde keine spezifische Arbeitshypothese formuliert, da die Fragestellung explorativ untersucht wurde (d. h. die Studie nicht auf die Beantwortung dieser Fragestellungen, sondern auf die primäre Fragestellung gepowert war).

Statistische Analyse: Analog zu den anderen Intention-to-Treat Analysen wurden für die therapeutische und technologische Allianz für die Outcome Parameter WAI-I und TAI-OT lineare Regressionen spezifiziert. Hierbei wurden nur die Patientinnen und Patienten der

Interventionsgruppen eingeschlossen, die mit der internetbasierten KVT-I zumindest begonnen hatten.

**4. Fragestellung:** Kam es in den drei Varianten der sektorenübergreifenden Versorgungsform zu negativen Effekten?

Arbeitshypothese: Für diese Fragestellung wurde keine spezifische Arbeitshypothese formuliert, da die Fragestellung explorativ untersucht wurde.

Statistische Analyse: Für den NEQ und die selbstentwickelte KVT-I spezifische Erweiterung NEQ Insomnia wurde der Mittelwert aller negativen Effekte sowie der Mittelwert der negativen Effekte, die spezifisch auf die Therapie zurückgeführt wurden, berechnet. Für die negativen Effekte, die auf die Therapie zurückgeführt wurden, wurde der Mittelwert für die dadurch erlebte Beeinträchtigung (Likert-Skala 1-5) berechnet. Darüber hinaus wurden negative Effekte berücksichtigt, falls Patientinnen oder Patienten diese an das Studienteam gemeldet haben.

## 2.7. Gesundheitsökonomische Evaluation

**Primäre Fragestellung:** Untersuchung der Kosten-Effektivität und des Kosten-Nutzwerts der neuen Versorgungsform im Vergleich zu TAU.

Arbeitshypothese: Es wird erwartet, dass die drei Varianten der sektorenübergreifenden neuen Versorgungsform zu einer signifikant höheren Anzahl an remittierten Patientinnen und Patienten bzw. einem signifikanten Gewinn an qualitätsadjustierten Lebensjahren (QALYs) und signifikant geringeren Kosten aus Sicht der gesetzlichen Krankenversicherung (GKV) wie aus gesamtgesellschaftlicher Perspektive im Vergleich zu TAU führen.

Statistische Analyse: Im Rahmen der gesundheitsökonomischen Analyse wurden die Kosten und die Anzahl an remittierten Patientinnen/Patienten (Kosten-Effektivitätsanalyse) sowie die Kosten und die QALYs (Kosten-Nutzwertanalyse) in der neuen Versorgungsform und der TAU-Gruppe verglichen. Dabei wurde zum einen die Sicht der GKV eingenommen und zum anderen eine gesamtgesellschaftliche Perspektive.

Die Kostenparameter für ambulante und stationäre Versorgung, ambulante und stationäre Rehabilitation, Arzneimittel sowie Heil- und Hilfsmittel wurden durch einen Selbstberichtsfragebogen erhoben und anschließend anhand standardisierter Bewertungssätze bewertet (Tabelle 17; Anhang 9 des Ergebnisberichts). Für die Bewertung von Arzneimittel wurde die Lauer-Taxe herangezogen. Zusätzlich wurden Ausgaben seitens der Patientinnen und Patienten sowie Produktivitätsverluste infolge von Arbeitsunfähigkeit und Präsentismus erfasst. Produktivitätsverluste aufgrund von Krankheitstagen wurden unter Verwendung des Humankapitalansatzes bewertet (Drummond et al., 2005), während durch Präsentismus verursachte Verluste anhand der Osterhaus-Methode quantifiziert wurden (Osterhaus et al., 2012). Fehlende Werte wurden mittels multipler Imputation (predictive mean matching) in Stata geschätzt (van Buuren & Groothuis-Oudshoorn, 2011). Es wurden Imputationsmodelle ohne Mehrebenenstruktur verwendet, da Mehrebenen-Imputationsmodelle zur Berücksichtigung der Cluster-Struktur keine Konvergenz erreichten, was vermutlich auf die geringe Anzahl an Clustern zurückzuführen war. Die Cluster-Zugehörigkeit wurde dennoch berücksichtigt, indem diese als Prädiktor in allen Imputationsmodellen aufgenommen wurde. Ebenso waren Geschlecht und Alter der Teilnehmenden sowie Baselinewerte Prädiktoren in allen Imputationsmodellen. Es wurden 10 Datensätze imputiert. Anschließend wurden die einzelnen Kostenkategorien zu Gesamtkosten zusammengefasst. Weiterhin wurden durchschnittliche Gesamtkosten aus GKV-Sicht und gesamtgesellschaftlicher Perspektive je Studiengruppe berechnet und mit Hilfe von unadjustierten Regressionsmodellen verglichen. Um die rechtsschiefe und nicht negative Kostenverteilung zu berücksichtigen, wurden anschließend an die Regressionsmodelle generalisierte lineare (GLM) Modelle der Gamma-Familie und robuste Standardfehler genutzt. Alter und Geschlecht dienten als sozio-demografische Kontrollvariablen. Die so generierten

Schätzwerte wurden als durchschnittlicher marginaler Effekt dargestellt, welcher den Kosteneffekt gemittelt über die beobachteten Kovariaten darstellt und eine direkte Interpretierbarkeit der Ergebnisse in Euro gewährleistet (Willan et al., 2004).

Neben der Kostenanalyse bildete die zuvor durchgeführte Effektivitätsanalyse die Grundlage für die Kosten-Effektivitätsanalyse. Zur Messung der gesundheitsbezogenen Lebensqualität für die Kosten-Nutzwertanalyse wurde der AQL-8D herangezogen und anschließend qualitätsadjustierte Lebensjahre (QALYs) berechnet (siehe Datenerhebung). Die Ergebnisse der gesundheitsökonomischen Evaluation wurden als inkrementelle Kosten-Effektivitäts-Relation abgebildet, welche die zusätzlich entstandenen Kosten in Euro im Verhältnis zu der zusätzlich zur Regelversorgung nachgewiesenen Wirksamkeit bzw. den zusätzlich nachgewiesenen Nutzwerten der neuen Versorgungsform darstellen. So wurden die inkrementellen Kosten pro gewonnenem QALY berechnet.

Basierend auf der Annahme, dass eine Korrelation zwischen dem Effektmaß und den Kosten besteht, wurde ein Mehrgleichungsmodell (engl. Seemingly unrelated regression models, SUR) herangezogen. Die statistische Unsicherheit der Schätzungen der Kosten-Effektivitäts-Relation wurde mit Hilfe der nicht-parametrischen Bootstrapping-Methode abgebildet. Darüber hinaus wurden Kosten-Effektivitäts-Akzeptanzkurven geschätzt, die die Akzeptanzwahrscheinlichkeit der neuen Versorgungsform in Abhängigkeit verschiedener Zahlungsbereitschaften darstellt.

**Sekundäre Fragestellung:** Untersuchung der Validität des TiC-P anhand von Routinedaten der BARMER.

**Arbeitshypothese:** Es handelt sich um eine explorative Untersuchung, daher wurden keine Hypothesen formuliert.

**Statistische Analyse:** Die Validität des TiC-P wurde durch die Bewertung der Übereinstimmung mit im Selbstbericht angegebenen Daten und durch die BARMER bereitgestellten Routinedaten für die ambulante Versorgung, Arzneimittel, stationäre Aufenthalte, Heil- und Hilfsmittel, ambulante und stationäre Rehabilitation sowie Arbeitsunfähigkeit bewertet, einschließlich des prozentualen absoluten Übereinstimmungsgrades, absoluter Unterschiede zwischen Daten aus dem Selbstbericht und Routinedaten und des Spearman-Rangkorrelationskoeffizienten ( $\rho$ ).

Die Erfassung, Verknüpfung und Auswertung von Primär- und BARMER-Routinedaten in der Studie erfolgte unter Berücksichtigung der gesetzlich festgelegten Datenschutzrichtlinien. Die Verwendung von Routinedaten im Rahmen des Projekts erfolgte auf der Grundlage von § 75 SGB X nach vorheriger Genehmigung der zuständigen Aufsichtsbehörde. Die Stelle für Gesundheitsökonomie an der Universität Ulm hatte keinen Zugriff auf pseudonymisierte Primär- wie personenbezogene Daten von Studienteilnehmenden. Zur Einhaltung des Datenschutzes wurde für das Datenlinkage ein weiteres Pseudonymisierungsverfahren eingesetzt, in dem eine Übermittlungs-ID erstellt wurde. Die Stelle für Gesundheitsökonomie erhielt jeweils den pseudonymisierten Primär- wie Routinedatensatz zur Datenverknüpfung anhand der Übermittlungs-ID. Nach einer Plausibilitätsüberprüfung der Daten durch die Stelle für Gesundheitsökonomie wurde die Kodierliste mit der Übermittlungs-ID vernichtet. Die Unabhängigkeit der Analysen wurde gewährleistet. Durch die Rekrutierung von Patientinnen und Patienten über Mailings der BARMER wurde sichergestellt, dass möglichst viele Projektteilnehmende bei der BARMER versichert waren.

### 3. Ergebnisse der Evaluation

#### 3.1. Erreichte Fallzahl und Dropouts

In die Studie wurden insgesamt 46 Hausärztinnen und Hausärzte sowie 333 Patientinnen und Patienten eingeschlossen (Abbildungen 2 und 3; Anhang 1 des Ergebnisberichts). Aufgrund des Projektendes wurde die Rekrutierung am 30. September 2022 beendet, obwohl die angestrebte Fallzahl bis zu diesem Zeitpunkt noch nicht erreicht wurde.

Ursprünglich wurde eine Stichprobengröße von 4268 Patientinnen und Patienten angestrebt. Aufgrund absehbarer Schwierigkeiten, dieses Rekrutierungsziel zu erreichen, wurde die Untersuchung der differentiellen Wirksamkeit aufgegeben und die neue erforderliche Stichprobengröße auf 1067 reduziert (Ausführliches Evaluationskonzept GET Sleep; Version 3.0, 16.12.2022). Diese neue Zielgröße wurde mit 333 tatsächlich rekrutierten Patientinnen und Patienten trotzdem deutlich unterschritten. Darüber hinaus haben nur 223 (67%) der 333 Patientinnen und Patienten die T3-Befragung abgeschlossen. Diese erhebliche Abweichung von der angestrebten Stichprobengröße sowie der zusätzliche Fragebogen-Dropout deuten auf ein deutliches Power-Problem hin. Infolgedessen sind die statistischen Analysen und die Verlässlichkeit der Studienergebnisse möglicherweise beeinträchtigt, da die Stichprobe nicht ausreichend groß sein könnte, um statistisch signifikante Effekte zu erfassen. Es ist daher notwendig, dies bei der Interpretation der Ergebnisse angemessen zu berücksichtigen.

#### 3.2. Baselinecharakteristika

Siebzig Prozent der Patientinnen und Patienten waren weiblich, das Durchschnittsalter lag bei 48,6 Jahren bei einer Standardabweichung (SD) von 13,2 Jahren. Der durchschnittliche ISI-Gesamtscore zum Zeitpunkt der Baseline-Messung betrug 18,6 Punkte (SD=4,2 Punkte). Weitere Baseline-Charakteristika der Teilnehmenden sind in Tabelle 1 dargestellt (Anhang 2 des Ergebnisberichts).

Von den 46 randomisierten Hausärztinnen und Hausärzten schlossen 18 (39,1%) auch tatsächlich Patientinnen und Patienten in die Studie ein. In die Intention-to-Treat-Analysen wurden die Daten von 333 Patientinnen und Patienten eingeschlossen, in die Per-Protocol-Analysen die Daten von 203 Patientinnen und Patienten. Das Flowchart der Patientinnen und Patienten (Abbildung 3, Anhang 1 des Ergebnisberichts) zeigt die Anzahl der Studienteilnehmenden je Studiengruppe.

#### 3.3. Inanspruchnahme der neuen Versorgungsform

Insgesamt haben in den IGs 134 Patientinnen und Patienten die hausärztliche Psychoedukation als erste Stufe des Stepped Care Modells in Anspruch genommen, während 134 Patientinnen und Patienten diese erste Stufe übersprungen haben (IG1: 84 Patientinnen und Patienten mit hausärztlicher Psychoedukation, 94 Patientinnen und Patienten mit Überspringen der Psychoedukation; IG2: 37 Patientinnen und Patienten mit hausärztlicher Psychoedukation, 24 Patientinnen und Patienten mit Überspringen der Psychoedukation; IG3: 13 Patientinnen und Patienten mit hausärztlicher Psychoedukation, 27 Patientinnen und Patienten mit Überspringen der Psychoedukation). Einen Überblick über die Inanspruchnahme der zweiten Stufe des Stepped Care Modells in den IGs gibt Abbildung 3 (Anhang 1 des Ergebnisberichts). Von den acht Einheiten der digitalen Intervention in der zweiten Stufe des Stepped Care Modells haben in den Interventionsgruppen 29 Patientinnen/Patienten eine Einheit abgeschlossen (IG1: 17; IG2: 7; IG3: 5), 10 Patientinnen/Patienten zwei Einheiten (IG1: 5; IG2: 3; IG3: 2), 7 Patientinnen/Patienten drei Einheiten (IG1: 6; IG2: 1; IG3: 0), 7 Patientinnen/Patienten vier Einheiten (IG1: 3; IG2: 3; IG3: 1), 14 Patientinnen/Patienten fünf Einheiten (IG1: 8; IG2: 2; IG3: 4), 12 Patientinnen/Patienten sechs Einheiten (IG1: 6; IG2: 4; IG3: 2), 6 Patientinnen/Patienten sieben Einheiten (IG1: 3; IG2: 2; IG3: 1) und 143 Patientinnen/Patienten alle acht Einheiten (IG1: 97; IG2: 28; IG3: 18). Gemäß Selbstausskunft der Patientinnen und Patienten im Rahmen der Befragungen zu T1, T2 und T3 wurden in Bezug auf die Stufe 3 des Stepped Care Modells zwei Patientinnen/Patienten

von ihren Hausärztinnen/Hausärzten in eine psychotherapeutische Fachbehandlung überwiesen (beide IG1), zwei Patientinnen/Patienten in eine psychiatrische Fachbehandlung (IG1: 1; IG2: 1) und drei Patientinnen/Patienten in eine schlafmedizinische Fachbehandlung (IG1: 2; IG2: 1).

### 3.4. Effektevaluation

**Primäre Fragestellung:** Untersuchung der Wirksamkeit der neuen Versorgungsform im Vergleich zu TAU im Hinblick auf den Schweregrad der Insomnie.

**Ergebnisse:** Die linearen Regressionsmodelle, die zur Untersuchung des Effektes der Studienbedingung auf den primären Outcome-Parameter ISI spezifiziert wurden (Intention-to-Treat-Analysen), sind in Tabelle 2 (Anhang 3.2 des Ergebnisberichts) abgebildet. In den paarweisen Gruppenvergleichen zeigten sich für den ISI keine signifikanten Unterschiede zwischen den Studienbedingungen (alle  $p > 0,1$ ). Tabelle 3 stellt die Per-Protocol-Analysen (Anhang 3.4 des Ergebnisberichts) dar. Auch bei den Per-Protocol-Analysen zeigten sich in den paarweisen Gruppenvergleichen keine signifikanten Gruppenunterschiede (alle  $p > 0,5$ ). In den Boxplots (Abbildungen 4 und 5; Anhang 3.1 und Anhang 3.3 des Ergebnisberichts) zeichnet sich der Trend ab, dass in allen Studienbedingungen eine Reduktion des ISI-Scores erfolgte. Die deskriptiven Daten weisen darauf hin, dass die Reduktion des ISI-Scores in den IGs stärker als in der TAU sein könnte.

Die Untersuchung der Responder mittels RCI zeigte, dass der Anteil der Responder in den IGs zwischen 72-75% lag, in der TAU-Bedingung bei 52% (Tabelle 4; Anhang 3.5 des Ergebnisberichts). Die logistische Regression, die zum Vergleich der Wahrscheinlichkeit einer Response zwischen den Studienbedingungen durchgeführt wurde, ist in Tabelle 5 (Anhang 3.5 des Ergebnisberichts) abgebildet. Es ergaben sich keine signifikanten Unterschiede zwischen den Studienbedingungen (alle  $p > 0,1$ ). Die Schätzungen der NNT zum Vergleich der IGs mit TAU lagen zwischen 7,3 bis 9,6 (Tabelle 6, Anhang 3.6 des Ergebnisberichts).

**Evaluation:** Die nicht-signifikanten Ergebnisse der Intention-to-Treat-Analyse sowie der weiteren Analysen des primären Outcome-Parameters ISI führen dazu, dass die Hypothese „Es wird erwartet, dass die drei Varianten der sektorenübergreifenden Versorgungsform zu einer signifikanten Verringerung der Insomnieschwere im Vergleich zu TAU führen.“ abgelehnt werden muss. Die Ablehnung dieser Hypothese begründet sich jedoch nicht unbedingt durch die Nicht-Überlegenheit der IGs im Vergleich zu TAU, sondern möglicherweise durch die nicht ausreichend gepowerte Studie. Deskriptiv deuten die Daten darauf hin, dass es einen Trend geben könnte, in dem es zu einer stärkeren Reduktion der Insomnieschwere in den IGs im Vergleich zu TAU geben könnte. Dies bedarf jedoch der Überprüfung im Rahmen einer konfirmatorischen, ausreichend gepowerten Studie.

**Sekundäre Fragestellung:** Untersuchung der Wirksamkeit der drei Varianten der sektorenübergreifende Versorgungsform bezogen auf die Schlafqualität, Lebensqualität, depressive Symptome, Angstsymptome, somatische Symptome sowie das Vorliegen einer depressiven Symptomatik nach der Behandlung.

**Ergebnisse:** Die linearen Regressionsmodelle, die zur Untersuchung des Effektes der Studienbedingung auf die sekundären Outcome-Parameter spezifiziert wurden (Intention-to-Treat-Analysen), sind in Tabelle 7 (Anhang 4.2 des Ergebnisberichts) abgebildet. In den paarweisen Gruppenvergleichen der Intention-to-Treat-Analysen zeigte sich lediglich für den Vergleich von IG2 vs. TAU für den sekundären Outcome-Parameter PSQI ein signifikanter Gruppenunterschied ( $\beta = -0,54$ ; 95% KI  $[-0,92, -0,16]$ ;  $p = 0,03$ ; alle anderen  $p > 0,1$ ). Tabelle 8 stellt die Per-Protocol Analysen dar (Anhang 4.4 des Ergebnisberichts). Bei den Per-Protocol-Analysen zeigten sich in den paarweisen Gruppenvergleichen keine signifikanten Unterschiede zwischen den Studienbedingungen für die sekundären Outcome-Parameter (alle  $p > 0,1$ ). Bei 13% der Patientinnen und Patienten der TAU-Gruppe lag zu T3 eine depressive Symptomatik (QIDS-SR16 $\geq$ 13) vor. Bei den IGs waren zwischen 7% bis 15% von einer

depressiven Symptomatik betroffen (Tabelle 9; Anhang 4.5 des Ergebnisberichts). Die Relative Risks der Poisson Regressionen mit robusten Standardfehlern, die zum Vergleich der IGs mit TAU durchgeführt wurden, lagen zwischen 0,57 und 1,15, erreichten jedoch keine statistische Signifikanz (alle  $p > 0,5$ ; Tabelle 10; Anhang 4.5 des Ergebnisberichts).

Evaluation: Die paarweisen Gruppenvergleiche, mit Ausnahme des paarweisen Gruppenvergleichs zwischen IG2 und TAU für das sekundäre Outcome Schlafqualität gemessen durch den PSQI, deuten darauf hin, dass keine Überlegenheit der IGs hinsichtlich der untersuchten sekundären Outcomes vorliegt. Es ist jedoch wichtig zu betonen, dass es sich um sekundäre Analysen handelt, die lediglich explorativ durchgeführt wurden. Darüber hinaus wurde die Studie nicht für die jeweiligen Outcomes gepowert. Daher wäre eine konfirmatorische Untersuchung für jeden spezifischen sekundären Outcome notwendig, um verlässliche Aussagen zu treffen.

**Moderations- und Mediationsanalysen:** Untersuchung potentieller Moderatoren und Mediatoren von Behandlungseffekten.

Ergebnisse: Tabelle 21 (Anhang 10 des Ergebnisberichts) zeigt die Mittelwerte und Standardabweichungen der ursprünglich angedachten Moderations- und Mediationsvariablen. Deskriptiv deuten die Daten darauf hin, dass die Summenwerte auf den einzelnen Skalen in den IGs gesenkt werden konnten.

Evaluation: Die deskriptiven Daten deuten darauf hin, dass in adäquat gepowerten Studien Mediations- und Moderationsanalysen mit diesen Variablen aufschlussreich sein könnten.

### 3.5. Prozessevaluation

Im Rahmen der Prozessevaluation wurden 39 Interviews mit Patientinnen und Patienten, sieben mit E-Coaches und sieben mit Hausärztinnen und Hausärzten geführt. In Bezug auf die hausärztliche Gruppe wurden vier Vor-Ort-Hausärztinnen und -Hausärzte und drei telemedizinisch arbeitende Hausärztinnen und Hausärzte interviewt. Eine Beschreibung des Kategoriensystems der qualitativen Interviews ist in Tabelle 12 (Anhang 5 des Ergebnisberichts) abgebildet.

**1. Fragestellung:** Welche Barrieren und Förderfaktoren bestehen für das neue Versorgungsmodell aus Sicht der Patientinnen und Patienten, E-Coaches und Hausärztinnen und Hausärzte?

Ergebnisse: Bezüglich der interpersonellen Qualität scheint die Mehrzahl der Patientinnen und Patienten eine Betreuung durch die E-Coaches geschätzt zu haben und es gibt Hinweise, dass die Betreuung zur Motivation und Adhärenz der Patientinnen und Patienten beigetragen haben könnte bzw. dass sich die Patientinnen und Patienten der IG2 und IG3 eine intensivere Betreuung durch die E-Coaches gewünscht hätten. Des Weiteren weisen Aussagen von sieben Patientinnen und Patienten und drei E-Coaches darauf hin, dass sich Patientinnen und Patienten ein weiteres Gespräch im Online-Training gewünscht hätten.

Hinsichtlich der technischen Qualität (z. B. inhaltliche Qualität der einzelnen Behandlungsstufen, Kompetenz der Behandlerinnen und Behandler) scheint es individuelle Unterschiede dahingehend zu geben, ob Patientinnen und Patienten von einer hausärztlichen Psychoedukation profitiert haben (je nachdem ob die Inhalte neu für die Patientinnen und Patienten waren/ sie diese Wiederholung schätzten). Die inhaltliche Qualität des Online-Trainings wurde von der Mehrheit der Patientinnen und Patienten positiv bewertet. Die Aussagen von sechs E-Coaches weisen darauf hin, dass es sich bei der Online-Intervention um eine für die Patientinnen und Patienten herausfordernde Intervention handele, bei der es auch zu Missverständnissen in der Umsetzung der Intervention kommen könne, weswegen eine Betreuung durch E-Coaches hilfreich sei. Bezüglich der dritten Stufe des Versorgungsmodells scheint es unzureichende Informationen gegeben zu haben, wie eine Einsteuerung in die dritte Stufe ablaufen könnte. Hinsichtlich auftretender negativer Effekte



wurden am häufigsten Erschöpfung, Konzentrationsschwierigkeiten und verminderte Produktivität berichtet.

Bezüglich der Umgebungsqualität wurden von den Patientinnen und Patienten diverse Aspekte des Online-Trainings genannt, die tendenziell positiv bewertet wurden (z. B. Videos und dazugehörige Transkripte, Wegbegleiterinnen und Wegbegleiter). Allerdings deuten Aussagen einiger Patientinnen und Patienten darauf hin, dass das Führen des Schlaftagebuchs eine Herausforderung war und acht Patientinnen/Patienten berichteten in diesem Kontext, dass es beim Führen des Schlaftagebuchs zu technischen Problemen gekommen sei.

Hinsichtlich der administrativen Qualität wurde der Ablauf von den meisten Patientinnen und Patienten als positiv bewertet. Sechs von ihnen berichten von langen Wartezeiten bis zum hausärztlichen Erstgespräch und zwei Patientinnen/Patienten berichteten von Verzögerungen beim Zugang zum Online-Training. Bezüglich der IG2 wurde deutlich, dass einigen Patientinnen und Patienten der Ablauf der Guidance-on-Demand nicht klar gewesen sei bzw. dass sie sich trotz Bedarf gehemmt gefühlt haben, sich Feedback von den Coaches einzufordern. Zudem wurde sowohl von Seiten der E-Coaches als auch der Hausärztinnen und Hausärzten thematisiert, dass ein Austausch zwischen diesen beiden Berufsgruppen gefehlt habe.

Evaluation: Die Auswertung der Interviews deutet darauf hin, dass aufgrund der häufig fehlenden Kommunikation zwischen Hausärztinnen/Hausärzten und E-Coaches sowie der fehlenden Informationen zur Einsteuerung in eine dritte Stufe die Behandlung in den IGs eher einem alleingestellten Online-Training als einem gestuften Versorgungsmodell entsprach. Bezüglich der ersten Stufe und auch bei Problemen mit der Einsteuerung in die dritte Stufe könnte eine Grundproblematik darin bestehen, dass es sich bei den meisten behandelnden Hausärztinnen und Hausärzten nicht um die persönlichen Hausärztinnen/Hausärzte der Patientinnen/Patienten handelte, sondern um Studienhausärztinnen und -hausärzte. Zudem scheint es unzureichend klar spezifizierte Strukturen für die Einsteuerung in die dritte Stufe gegeben zu haben.

In Bezug auf das Online-Training zeigte die Auswertung ein gemischtes Meinungsbild der Patientinnen und Patienten, wie es bei internet- und mobilbasierten Interventionen häufig berichtet wird. Aus qualitativer Perspektive scheint eine intensive Begleitung des Online-Trainings, wie sie in der IG1 umgesetzt wurde, sowohl von den meisten Patientinnen und Patienten als auch von den E-Coaches bevorzugt zu werden. Es ergaben sich Hinweise darauf, dass bei einer "Guidance-on-Demand", wie sie die IG2 vorsah, Patientinnen und Patienten klarere Hinweise zur Nutzung dieses Format gegeben werden müsste, um Unsicherheiten und einer Nicht-Inanspruchnahme im Bedarfsfall vorzubeugen. Es gilt zu beachten, dass die identifizierten Themen auf den Aussagen einzelner Personen beruhen. Um die Repräsentativität dieser Aussagen zu bestimmen, wäre eine strukturierte Klärung erforderlich. Die detaillierten Ergebnisse der qualitativen Untersuchung können Anhang 5 des Ergebnisberichts entnommen werden.

## **2. Fragestellung: Wie ist die patientenseitige Zufriedenheit mit der Behandlung?**

Ergebnisse: In den paarweisen Gruppenvergleichen, die zum Vergleich des CSQ-I zwischen den Studienbedingungen spezifiziert wurden, zeigten sich keine signifikanten Gruppenunterschiede (alle  $p > 0,5$ ; Tabelle 13; Anhang 6 des Ergebnisberichts). In den Boxplots (Abbildung 8; Anhang 6 des Ergebnisberichts) zeichnet sich deskriptiv der Trend ab, dass der CSQ-I Gesamtscore in der TAU im Vergleich zu den IGs geringer war.

Evaluation: Die paarweisen Gruppenvergleiche deuten nicht auf eine Überlegenheit der IGs hinsichtlich der patientenseitigen Zufriedenheit hin. Es handelt sich hierbei um sekundäre Analysen, die lediglich explorativ durchgeführt wurden und die vorliegende Studie war für den Outcome patientenseitige Zufriedenheit nicht gepowert. Daher wäre eine confirmatorische Untersuchung notwendig, um verlässliche Aussagen zu treffen.

**3. Fragestellung:** Wie ist die therapeutische Allianz zu den E-Coaches und die technologische Allianz zur internetbasierten KVT-I?

Ergebnisse: In den linearen Regressionen für den Outcome-Parameter WAI-I zeigten sich signifikante Unterschiede zwischen IG1 und IG2 ( $\beta=0,50$ ; 95% KI [0,14, 0,86];  $p=0,02$ ) sowie zwischen IG1 und IG3 ( $\beta=0,48$ ; 95% KI [0,06, 0,89];  $p=0,05$ ). Die weiteren paarweisen Vergleiche für die Outcomes WAI-I und TAI-OT zeigten keine signifikanten Unterschiede (alle  $p>0,05$ ; Tabelle 14; Anhang 7 des Ergebnisberichts).

Evaluation: Die paarweisen Gruppenvergleiche deuten darauf hin, dass die therapeutische Allianz in der IG1 stärker als in den anderen beiden IGs ausgeprägt sein könnte. Bezüglich der technologischen Allianz ergaben sich bei den paarweisen Gruppenvergleichen keine Hinweise für Unterschiede zwischen den IGs. Auch hier gilt es zu beachten, dass es sich um explorative sekundäre Analysen handelt, die konfirmatorisch belegt werden müssten.

**4. Fragestellung:** Kam es in den drei Varianten der sektorenübergreifenden Versorgungsform zu negativen Effekten?

Ergebnisse: Die mittlere Anzahl der negativen Effekte und die entsprechende Anzahl der negativen Effekte, die auf die Therapie attribuiert wurden, sind in Tabelle 15 und 16 abgebildet (Anhang 8 des Ergebnisberichts). In den IGs lag der Mittelwert für durch den NEQ erhobene negative Effekte zwischen 2,0 und 2,8 und bei der KVT-I spezifischen Erweiterung NEQ Insomnia zwischen 4,3 und 5,7. Sechs Patientinnen und Patienten berichteten von Unfällen während des Interventionszeitraum, wovon eine Person einen Treppensturz auf die Intervention zurückführte. Darüber hinaus kam es zu dem Widerruf der Studienteilnahme einer Person aufgrund eines Unfalls an einer Kreissäge.

Evaluation: Es gibt Hinweise, dass die Bettzeitrestriktion, die ein fester Bestandteil der KVT-I ist, ein Risiko für negative Effekte birgt (Kyle et al., 2011). Im Vergleich zu anderen Studien war das Ausmaß der negativen Effekte in der vorliegenden Studie in einem erwartbaren Ausmaß; dennoch ist zu beachten, dass die Forschungslage zu den negativen Effekten der KVT-I aktuell unzureichend ist. Das Auftreten von negativen Effekten und insbesondere von Unfällen unterstreicht die Notwendigkeit einer sorgfältigen Aufklärung über die KVT-I als Therapieoption.

### 3.6. Gesundheitsökonomische Evaluation

**Fragestellung:** Untersuchung der Kosten-Effektivität und des Kosten-Nutzwerts der neuen Versorgungsform im Vergleich zu TAU.

Ergebnisse: Es zeigten sich keine Unterschiede in der Anzahl an remittierten Patientinnen und Patienten ( $ISI\text{-Score}\leq 7$ ) zwischen den Studienbedingungen. Die Remissionsrate variierte zwischen 24% in der TAU-Gruppe und 33% in IG3. Es gab keine signifikanten Unterschiede im QALY-Gewinn zwischen den Studienbedingungen (TAU-Gruppe: 0,33; IG1: 0,33; IG2: 0,33; IG3: 0,34;  $p=1,00$ ; Tabelle 18; Anhang 9 des Ergebnisberichts).

Die Kosten pro Patientin/Patient für die neue Versorgungsform betragen: Psychoedukation und ein weiterer Hausarztbesuch: 51,60€, IG1: 490,72€, IG2: 221,92€, IG3: 132,32€, und GET.ON Arztbrief: 8,34€. Die TAU-Gruppe hatte aus Sicht der GKV die niedrigsten durchschnittlichen Kosten im Nachbeobachtungszeitraum von sechs Monaten (494€, 95% KI: 142€-846€). Die unadjustierten Regressionsmodelle zeigten keinen signifikanten Kostenunterschied zwischen der TAU-Behandlung und IG2 (89€,  $p=0,66$ ) sowie IG3 (100€,  $p=0,76$ ), während IG1 (1.377€,  $p<0,001$ ) signifikant höhere Kosten verursachte, hauptsächlich durch vermehrte stationäre Aufenthalte. Von den Patientinnen/Patienten berichteten 3% (9/333) im Follow-Up über stationäre Aufenthalte, hauptsächlich in IG1 ( $n=6$ ; Tabelle 19; Anhang 9 des Ergebnisberichts). Die per Mittelwertsdifferenz analysierten Kostenunterschiede wurden durch adjustierte GLM-Modelle bestätigt. Der marginale Kostenunterschied zwischen Standardbehandlung und IG1 (1.458€,  $p<0,001$ ) sowie IG2 (250€,  $p=0,02$ ) war quantitativ

höher, während er für IG3 (98€,  $p=0,33$ ) unverändert blieb (Tabelle 20; Anhang 9 des Ergebnisberichts). Aus gesamtgesellschaftlicher Perspektive verursachte ebenfalls die TAU-Behandlung die geringsten Kosten im Nachbeobachtungs-zeitraum (3.567€, 95% KI: 1981€-5153€). Der Unterschied zwischen TAU-Behandlung und IG1 war signifikant (2.497€,  $p=0,02$ ), während er für IG2 (592€,  $p=0,64$ ) und IG3 (2.830€,  $p=0,30$ ) nicht signifikant war. Die adjustierten GLM-Modelle bestätigten diese Ergebnisse (Tabelle 20; Anhang 9 des Ergebnisberichts).

Eine umfassende Kosten-Effektivitäts-Analyse sowie eine Kosten-Nutzwert-Analyse wurden nicht durchgeführt, da keine signifikanten Unterschiede in der Effektivität zwischen der neuen Versorgungsform und der TAU-Behandlung nachweisbar waren und die Kosten für die neue Versorgungsform aus beiden Perspektiven höher waren.

**Evaluation:** Die gesundheitsökonomische Analyse ergab keine statistisch signifikante Reduktion der durchschnittlichen Kosten durch die neue Versorgungsform im Vergleich zur TAU-Behandlung, weder aus der Perspektive der GKV noch aus gesamtgesellschaftlicher Sicht. Zusätzlich konnten keine signifikanten Unterschiede in den klinischen Effekten zwischen den Studiengruppen nachgewiesen werden. Die Nichtannahme der Hypothese bezüglich der klinischen Überlegenheit und Kostenreduktion der neuen Versorgungsform im Vergleich zur TAU-Behandlung könnte möglicherweise auf die unzureichende Power der Studie zurückzuführen sein.

**Sekundäre Fragestellung:** Untersuchung der Validität des TiC-P anhand von Routinedaten der BARMER.

**Ergebnisse:** Von den 333 Teilnehmenden der Studie wurden Routinedaten von 166 Teilnehmenden durch die BARMER bereitgestellt. Über die Baseline- und Follow-Up-Erhebungen hinweg entsprach dies insgesamt 402 individuellen Selbstangaben. Nach Überprüfung der Routinedaten wurde von einem Abgleich mit Kontakten zu Fachärztinnen und Fachärzten für Hals-Nasen-Ohrenheilkunde, Innere Medizin und Pneumologie, Nervenheilkunde, Neurologie und Psychiatrie, Psychiatrie und Psychotherapie, psychosomatische Medizin und Psychotherapie sowie psychologische Psychotherapeutinnen/Psychotherapeuten, stationären Krankenhausaufnahmen, Rehabilitationsmaßnahmen (stationär und ambulant), Arzneimitteln, sowie Heil- und Hilfsmittel aufgrund der insgesamt geringen Inanspruchnahme abgesehen. Somit wurde die Validität des TiC-P anhand von Facharztkontakten in der Allgemeinmedizin und der Anzahl an Arbeitsunfähigkeitstagen überprüft. Die selbstberichtete Anzahl von Kontakten mit einem Facharzt/Fachärztin für Allgemeinmedizin in den letzten drei Monaten betrug durchschnittlich 1,5 (Standardabweichung, SD 1,8) und erstreckte sich von 0 bis 10. Diese Angaben wiesen eine hohe Korrelation von  $\rho = 0,77$  mit den entsprechenden Routinedaten auf. Die absolute Übereinstimmung zwischen Selbstbericht und Routinedaten bezüglich der Kontakte mit einem Facharzt/Fachärztin für Allgemeinmedizin betrug 71,1%. Die mittlere Differenz zwischen den Selbstangaben und den Routinedaten belief sich auf 0,1 (SD=0,9). Selbstbericht und Routinedaten zu AU-Tagen lagen von 122 Teilnehmenden vor. AU-Tage in den letzten drei Monaten betrug durchschnittlich 7,7 Tage (SD=20,5) und erstreckte sich von 0 bis 90 Tage. Die absolute Übereinstimmung zwischen Selbstbericht und Routinedaten betrug 63,9%. Die Differenz zwischen Selbstbericht und Routinedaten bzgl. der Arbeitsunfähigkeit betrug durchschnittlich 1,7 Tage (SD=9,8 Tage). Die Korrelation zwischen selbstberichteten AU-Tagen und Routinedaten betrug 0,79.

**Evaluation:** Die Übereinstimmung zwischen Selbstbericht und Routinedaten über die Anzahl der Kontakte mit Fachärztinnen/Fachärzten der Allgemeinmedizin bzw. AU-Tage war zufriedenstellend, was auf eine akzeptable Validität dieser beiden Items hinweist. Aufgrund geringer Inanspruchnahme war es nicht möglich, weitere Ressourcennutzung im Gesundheitswesen zu untersuchen. Daher ist weitere Forschung erforderlich, um die Validität des TiC-Ps bzgl. weiterer Gesundheitsversorgung zu bewerten.

#### **4. Schlussfolgerungen und Empfehlungen des Evaluators**

Die vorliegenden Ergebnisse erlauben keine Empfehlung für die Integration der neuen Versorgungsform in die Regelversorgung. Diese Schlussfolgerung stützt sich sowohl auf die Ergebnisse der Effektevaluation als auch die Ergebnisse der Prozessevaluation. In der Effektevaluation konnte eine überlegene Wirksamkeit der neuen Versorgungsform gegenüber TAU weder im Hinblick auf das primäre Outcome, die Insomnieschwere, noch bezüglich der untersuchten Sekundäroucomes nachgewiesen werden. Es ist jedoch wichtig zu betonen, dass die Ergebnisse möglicherweise durch die geringe Stichprobengröße und die nicht ausreichende statistischen Power der Studie erklärbar sind. Deskriptiv legen die Daten den Trend einer stärkeren Reduktion der Insomnieschwere in den IGs im Vergleich TAU nahe. Diese Beobachtung erfordert jedoch eine Überprüfung in einer confirmatorischen Studie mit ausreichender Power. Neben den Ergebnissen der Prozessevaluation sollten dabei auch die Rekrutierungsschwierigkeiten dieser Studie berücksichtigt werden. Die ursprünglich geplanten Moderations- und Mediationsanalysen konnten aufgrund der geringen Stichprobengröße sowie des Fehlens von Gruppenunterschieden in der primären Analyse nicht durchgeführt werden.

Die Schwierigkeiten bei der Rekrutierung von Vor-Ort-Hausärztinnen und -Hausärzten unterstreichen die Notwendigkeit einer gründlichen Vorbereitung vor der Implementierung eines gestuften Versorgungsmodells. Zusammen mit den Ergebnissen der qualitativen Studie, die darauf hinweisen, dass nur ein Teil der Patientinnen/Patienten einen Mehrwert in der hausärztlichen Erstbehandlung sah, stellt sich die Frage, ob eine hausärztliche Erstbehandlung als angemessene Einstiegsstufe für ein solches Versorgungsmodell geeignet ist. Die Ergebnisse der qualitativen Studie deuten darauf hin, dass eine weitere Optimierung der intersektoralen Kommunikation zwischen Hausärztinnen/Hausärzten und E-Coaches erforderlich wäre. Darüber hinaus verdeutlichen die Ergebnisse der qualitativen Studie die Notwendigkeit klarerer Strukturen für die Einsteuerung in die dritte Stufe des Versorgungsmodells.

Während die quantitative Evaluation keine signifikanten Unterschiede in der Wirksamkeit der drei Varianten der therapeutischen Begleitung im Online-Training zeigen konnte, deuten die Ergebnisse der qualitativen Studie daraufhin, dass während die Bedürfnisse hinsichtlich des Ausmaßes der therapeutischen Begleitung stark individuell variieren, die meisten Patientinnen und Patienten eine intensivere therapeutische Begleitung bevorzugen würden. Aufgrund der geringen Stichprobengröße wurde in den quantitativen Analysen keine Unterscheidung der Stichprobe in vier Gruppen (i.e., Insomnie ohne psychische und körperliche Komorbidität, Insomnie mit psychischer Komorbidität, Insomnie mit körperlicher Komorbidität, Insomnie mit psychischer und körperlicher Komorbidität) vorgenommen. Zukünftige Studien sollten untersuchen, ob verschiedene Gruppen von Patientinnen und Patienten eine unterschiedliche Intensität der therapeutischen Begleitung bedürfen.

Die vorliegenden Ergebnisse unterstreichen die Herausforderungen, die bei der Implementierung neuer Versorgungsformen entstehen können. Insbesondere verdeutlichen die Rekrutierungsschwierigkeiten und Probleme in der intersektoralen Kommunikation, dass Strukturen für neue Versorgungsmodelle vor ihrer Einführung eingehend entwickelt und in Pilotprojekten erprobt werden sollten.

#### **5. Literaturverzeichnis**

Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5:233-237.

Akronym: GET Sleep

Förderkennzeichen: 01NVF18030

Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med* 2001;2:297-307.

Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, Ebert DD. Reliability and validity of assessing user satisfaction with web-based health interventions. *J Med Internet Res* 2016;18:e234.

Bouwman C, de Jong K, Timman R, Zijlstra-Vlasveld M, van der Feltz-Cornelis C, Tan Swan S, Hakkaart-van Roijen L. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). *BMC Health Serv Res* 2013;13:217.

Buth S, Holzbach R, Martens MS, Neumann-Runde E, Meiners O, Verthein U. Problematic medication with benzodiazepines, "Z-drugs", and opioid analgesics. *Dtsch Arztebl Int* 2019;116:607-614.

Buntrock C, Lehr D, Smit F, Horvath H, Berking M, Spiegelhalter K, Riper H, Ebert DD. Guided internet-based cognitive behavioral therapy for insomnia: health-economic evaluation from the societal and public health care perspective alongside a randomized controlled trial. *J Med Internet Res* 2021;23:e25609.

Buyse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989;28:193-213.

Carpenter JS, Andrykowski MA. Psychometric evaluation of the Pittsburgh Sleep Quality Index. *J Psychosom Res* 1998;45:5-13.

Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983;24:385-396.

Dagger TS, Sweeney JC, Johnson LW. A hierarchical model of health service quality: scale development and investigation of an integrated model. *J Serv Res* 2007;10: 123-142.

Drummond M, Sculpher M, Torrance G, O'Brian B, Stoddart G. Methods for the economic evaluation of health care programmes (3<sup>rd</sup> ed). Oxford University Press, 2005.

Espie CA, Inglis SJ, Harvey L, Tessier S. Insomniacs' attributions. Psychometric properties of the dysfunctional beliefs and attitudes about sleep scale and the sleep disturbance questionnaire. *J Psychosom Res* 2000;48:141-148.

Forsell E, Jernelöv S, Blom K, Kraepelien M, Svanborg C, Andersson G, Lindefors N, Kaldo V. Proof of concept for an adaptive treatment strategy to prevent failures in internet-delivered CBT: a single-blind randomized clinical trial with insomnia patients. *Am J Psychiatry* 2019;176:315-323.

Gagnon C, Bélanger L, Ivers H, Morin CM. Validation of the Insomnia Severity Index in primary care. *J Am Board Fam Med* 2013;26:701-710.

Gao A, Ge L, Liu M, Niu M, Chen Y, Sun Y, Chen J, Yao L, Wang Q, Li Z, Xu J, Li M, Hou L, Shi J, Yang K, Cai Y, Li L, Zhang J, Tian J. Comparative efficacy and acceptability of cognitive behavioral therapy delivery formats for insomnia in adults: A systematic review and network meta-analysis. *Sleep Med Rev* 2022;64:101648.

Garnefski N, Kraaij V. Cognitive emotion regulation questionnaire – development of a short 18-item version (CERQ-short). *Pers Individ Dif* 2006;41:1045-1053.

Gierk B, Kohlmann S, Kroenke K, Spangenberg L, Zenger M, Brähler E, Löwe B. The somatic symptom scale-8 (SSS-8): a brief measure of somatic symptom burden. *JAMA Intern Med* 2014;174:399-407.

Gieselmann A, de Jong-Meyer R, Pietrowsky R. Kognitive und körperliche Erregung in der Phase vor dem Einschlafen. Die deutsche Version der Pre-Sleep Arousal Scale (PSAS). *Z Klin Psychol Psychother* 2012;41:73-80.

Akronym: GET Sleep

Förderkennzeichen: 01NVF18030

Gómez Penedo JM, Berger T, Grosse Holtforth M, Krieger T, Schröder J, Hohagen F, Meyer B, Moritz S, Klein JP. The Working Alliance Inventory for guided Internet interventions (WAI-I). *J Clin Psychol* 2020;76:973-986.

Hakkaart-van Roijen L, van Straten A, Donker M, Tiemens B. Manual Trimbos/iMTA questionnaire for costs associated with psychiatric illness (TiC-P). Erasmus University, 2002.

Hasan F, Tu YK, Yang CM, Gordon CJ, Wu D, Lee HC, Yuliana LT, Herawati L, Chen TJ, Chiu HY. Comparative efficacy of digital cognitive behavioral therapy for insomnia: a systematic review and network meta-analysis. *Sleep Med Rev* 2022;61:101567.

Helfferrich C. Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews. VS Verlag für Sozialwissenschaften Wiesbaden, 2011.

Herrero R, Vara MD, Miragall M, Botella C, Garcia-Palacios A, Riper H, Kleiboer A, Banos RM. Working Alliance Inventory for Online Interventions-Short Form (WAI-TECH-SF): the role of the therapeutic alliance between patient and online program in therapeutic outcomes. *Int J Environ Res Public Health* 2020;17:6169.

Hertenstein E, Feige B, Gmeiner T, Kienzler C, Spiegelhalder K, Johann A, Jansson-Fröjmark M, Palagini L, Rücker G, Riemann D, Baglioni C. Insomnia as a predictor of mental disorders: A systematic review and meta-analysis. *Sleep Medicine Reviews* 2019;43:96-105.

Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991;59:12-19.

Kuckartz U, Rädiker S. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Beltz Juventa, 2022.

Kyle SD, Morgan K, Espie CA. Insomnia and health-related quality of life. *Sleep Med Rev* 2010;14:69-82.

Kyle SD, Morgan K, Spiegelhalder K, Espie CA. No pain, no gain: an exploratory within-subjects mixed-methods evaluation of the patient experience of sleep restriction therapy (SRT) for insomnia. *Sleep Med* 2011;12:735-747.

Lamoureux BE, Linardatos E, Fresco DM, Bartko D, Logue E, Milo L. Using the QIDS-SR16 to identify major depressive disorder in primary care medical patients. *Behav Ther* 2010;41:423-431.

Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *New Engl J Med* 1988;318:1728-1733.

Li M, Zhang XW, Hou WS, Tang ZY. Insomnia and risk of cardiovascular disease: a meta-analysis of cohort studies. *Int J Cardiol* 2014;176:1044-1047.

Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, Herzberg PY. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Med Care* 2008;46:266-274.

Mastin DF, Bryson J, Corwyn R. Assessment of sleep hygiene using the sleep hygiene index. *J Behav Med* 2006;29:223-227.

Mendoza TR, Wang XS, Cleeland CS, Morrissey M, Johnson BA, Wendt JK, Huber SL. The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory. *Cancer* 1999;85:1186-1196.

Morin CM, Stone J, Trinkle D, Mercer J, Remsberg S. Dysfunctional beliefs and attitudes about sleep among older adults with and without insomnia complaints. *Psychol Aging* 1993;8:463-467.

Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep* 2011;34:601-608.

Akronym: GET Sleep

Förderkennzeichen: 01NVF18030

Morin CM, Drake CL, Harvey AG, Krystal AD, Manber R, Riemann D, Spiegelhalder K. Insomnia disorder. *Nat Rev Dis Primers* 2015;1:15026.

Nicassio PM, Mendlowitz DR, Fussell JJ, Petras L. The phenomenology of the pre-sleep state: the development of the pre-sleep arousal scale. *Behav Res Ther* 1985;23:263-271.

Osterhaus JT, Gutterman DL, Plachetka JR. Healthcare resource and lost labour costs of migraine headache in the US. *Pharmacoeconomics* 1992;2:67-76.

Qaseem A, Kansagara D, Forcica MA, Cooke M, Denberg TD. Management of chronic insomnia disorder in adults: a clinical practice guideline from the American College of Physicians. *Ann Intern Med* 2016;165:125-133.

Patton MQ. *Qualitative Research & Evaluation Methods* (3rd edition). Thousand Oaks: Sage, 2002.

Richardson J, Iezzi A, Khan MA, Maxwell A. Validity and reliability of the Assessment of Quality of Life (AQoL)-8D multi-attribute utility instrument. *Patient* 2014;7:85-96.

Riemann D, Baum E, Cohrs S, Crönlein T, Hajak G, Hertenstein E, Kloese P, Langhorst J, Mayer G, Nissen C, Pollmächer T, Rabstein S, Schlarb A, Sitter H, Weess HG, Wetter T, Spiegelhalder K. S3-Leitlinie Nicht erholsamer Schlaf/ Schlafstörungen – Kapitel Insomnie bei Erwachsenen, Update 2016. *Somnologie* 2017;21:2-44.

Riemann D, Espie CA, Altena E, Arnardottir ES, Baglioni C, Bassetti CLA, Bastien C, Berzina N, Bjorvatn B, Dikeos D, Dolenc Groselj L, Ellis JG, Garcia-Borreguero D, Geoffroy PA, Gjerstad M, Goncalves M, Hertenstein E, Hoedlmoser K, Hion T, Holzinger B, Janku K, Jansson-Fröjmark M, Järnefelt H, Jernelöv S, Jennum PJ, Khachatryan S, Krone L, Kyle SD, Lancee J, Leger D, Lupusor A, Ruivo D, Marques DR, Nissen C, Palagini L, Paunio T, Perogamvros L, Pevernagie D, Schabus M, Shochat T, Szentkiralyi A, van Someren E, van Straten A, Wichniak A, Spiegelhalder K. The European Insomnia Guideline: an update on the diagnosis and treatment of insomnia 2023. *J Sleep Res* 2023;32:e14035.

Rozenal A, Kottorp A, Forsström D, Månsson K, Boettcher J, Andersson G, Furmark T, Carlbring P. The Negative Effects Questionnaire: psychometric properties of an instrument for assessing negative effects in psychological treatments. *Behav Cogn Psychother* 2019;47:559-572.

Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 2003;54:573-583.

Salomonsson S, Santoft F, Lindsäter E, Ejeby K, Ljótsson B, Öst LG, Ingvar M, Lekander M, Lagerlöf EH. Stepped care in primary care - guided self-help and face-to-face cognitive behavioural therapy for common mental disorders: a randomized controlled trial. *Psychol Med* 2018;48:1644-1654.

Sandelowski M. Sample size in qualitative research. *Res Nurs Health* 1995;18:179-183.

Schlack R, Hapke U, Maske U, Busch M, Cohrs S. Häufigkeit und Verteilung von Schlafproblemen und Insomnie in der deutschen Erwachsenenbevölkerung: Ergebnisse der Studie zur Gesundheit Erwachsener in Deutschland (DEGS1). *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2013;56:740-748.

Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006;166:1092-1097.

Thiart H, Ebert DD, Lehr D, Nobis S, Buntrock C, Berking M, Smit F, Riper H. Internet-based cognitive behavioral therapy for insomnia: a health economic evaluation. *Sleep* 2016;39:1769-1778.

Akronym: GET Sleep

Förderkennzeichen: 01NVF18030

Trujols J, Diego-Adeliño J, Feliu-Soler A, Iraurgi I, Puigdemont D, Álvarez E, Pérez V, Portella MJ. Looking into the effect of multi-item symptom domains on psychometric characteristics of the Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR16). *Psychiatry Res* 2018;267:126-130.

van Buuren S, Groothuis-Oudshoorn K. Mice: Multivariate imputation by chained equations in R. *J Stat Softw* 2011;45:1-67.

van Straten A, van der Zweerde T, Kleiboer A, Cuijpers P, Morin CM, Lancee J. Cognitive and behavioral therapies in the treatment of insomnia: a meta-analysis. *Sleep Med Rev* 2018;38:3-16.

VERBI Software. MAXQDA 2022. Version 22.8.0. Berlin: VERBI Software, 2021. Online verfügbar unter [maxqda.com](http://maxqda.com).

Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Econ* 2004;13:461-475.

Zhou ES, Michaud AL, Recklitis CJ. Developing efficient and effective behavioral treatment for insomnia in cancer survivors: results of a stepped care trial. *Cancer* 2020;126:165-173.

Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004;159:702-706.

## **6. Anhang**

Nicht vorhanden.

## **7. Anlagen**

Nicht vorhanden.