Psychometric properties of sleep questionnaires investigating sleep quality in older people: a systematic review

Farkad Ibrahim, KULeuven

Levi Van Hootegem, KULeuven

Promotor: Nicolas Delvaux, KULeuven

Master of Family Medicine

Masterproef Huisartsgeneeskunde
Academiejaar: [2021 – 2022]

Deze masterproef is een examendocument dat niet werd gecorrigeerd voor eventueel vastgestelde fouten. Zonder voorafgaande schriftelijke toestemming van zowel de promotor(en) als de auteur(s) is overnemen, kopiëren, gebruiken of realiseren van deze uitgave of gedeelten ervan verboden. Voor aanvragen tot of informatie i.v.m. het overnemen en/of gebruik en/of realisatie van gedeelten uit deze publicatie, wendt u tot de universiteit waaraan de auteur is ingeschreven.

Voorafgaande schriftelijke toestemming van de promotor(en) is eveneens vereist voor het aanwenden van de in dit afstudeerwerk beschreven (originele) methoden, producten, schakelingen en programma's voor industrieel of commercieel nut en voor de inzending van deze publicatie ter deelname aan wetenschappelijke prijzen of wedstrijden.

Abstract

Background: In the older GP population insomnia is a frequent underdiagnosed complaint. Many questionnaires or scales exist to explore insomnia. Unfortunately, no questionnaire is widely recognized as a screening tool for a general practitioner.
Objective: This systematic review aimed to identify and evaluate the psychometric properties of questionnaires and scales used to evaluate sleep quality in people aged 65 and above. Moreover, to summarize which questionnaire to use as a diagnostic tool for elderly.
Methods: PubMed was searched from 1 March 2012 to 30 June 2021 for studies published in English and Dutch. A systematic literature search was conducted by two independent researchers. The quality assessment was conducted using the Quadas-2 tool. Reliability of the studies was reviewed by assessing Cronbach's α. Validity was evaluated of all included articles by analysing content, construct or criterion validity. To determine usefulness each sleep questionnaire as screening tool sensitivity and specificity was evaluated.
Results: Only 4 new articles after previous review that were published met inclusion and exclusion criteria. Athens Insomnia Scale (AIS), Insomnia Severity Index (ISI), PROMIS Sleep Disturbance Scale, Richards-Campbell Sleep Questionnaire (RCSQ-C) and Pittsburgh Sleep Quality Index were studied for reliability, validity. They showed all adequate internal consistency. The sleep questionnaires were validated in the elderly population. All new sleep questionnaires need more investigation for use as a diagnostic screening tool.
Conclusion: Most sleep questionnaires lack investigation for use as a screening tool. The Insomnia in the Elderly Scale and the Minimal Insomnia Symptom Scale are most promising but need more reproducibility to confirm current results.

Abstract

Achtergrond: Bij de oudere patiënten populatie is slapeloosheid een onder gediagnosticeerde klacht. Er zijn veel vragenlijsten of schalen om slapeloosheid te onderzoeken. Er is echter geen enkele vragenlijst die als een diagnostische tool gebruikt kan worden in de huisartsen praktijk.
Doel: Deze systematische review was gericht op het identificeren en evalueren van de psychometrische eigenschappen van vragenlijsten en schalen die worden gebruikt om de slaapkwaliteit te evalueren bij mensen van 65 jaar en ouder. Bovendien, om te evalueren welke vragenlijst kunnen gebruikt worden als diagnostisch hulpmiddel bij ouderen.
Methode: Van 1 maart 2012 tot 30 juni 2021 is in PubMed gezocht naar studies gepubliceerd in het Engels en Nederlands. Een systematisch literatuuronderzoek werd uitgevoerd door twee onafhankelijke onderzoekers. De kwaliteitsbeoordeling is uitgevoerd met behulp van de tool Quadas-2. De betrouwbaarheid van de onderzoeken werd beoordeeld door Cronbach's te beoordelen. Validiteit werd geëvalueerd van alle opgenomen artikelen door inhouds-, construct- of criteriumvaliditeit te analyseren. Om de bruikbaarheid van elke slaapvragenlijst als screeningsinstrument te bepalen, werd de gevoeligheid en specificiteit geëvalueerd.
Resultaten: Slechts 4 nieuwe artikelen werden gevonden die antwoorden aan de inclusie en exclusiecriteria. Athens Insomnia Scale (AIS), Insomnia Severity Index (ISI), PROMIS Sleep Disturbance Scale, Richards-Campbell Sleep Questionnaire (RCSQ-C) en Pittsburgh Sleep Quality Index werden onderzocht op betrouwbaarheid en validiteit. Ze toonden allemaal voldoende interne consistentie. De slaapvragenlijsten werden gevalideerd bij de oudere populatie. Alle nieuwe slaapvragenlijsten hebben meer onderzoek nodig voor gebruik als diagnostisch screeningsinstrument.
Conclusie: Bij de meeste slaapvragenlijsten ontbreekt een onderzoek voor gebruik als screeningsinstrument. De Insomnia in the Elderly Scale en de Minimal Insomnia Symptom Scale zijn het meest veelbelovend, maar hebben meer reproduceerbaarheid nodig om de huidige resultaten te bevestigen.

Introduction

Epidemiological studies estimate that by 2050, 22% of the population will be above 60 years(1). Research shows that about 50% of the old population will develop insomnia (1) (2)(3). However, the cause of this high incidence rate is not well identified. Age-related changes in the circadian rhythm of sleep, the partial decline in functionality of sleep control centers, and the presence of preexisting underlying comorbidities are currently the most accepted explanations (4). This increase in incidence of insomnia in an elderly population is an important health risk because insomnia is associated with increased all-cause mortality (2)(3).

The normal sleep cycle consists of four stages. The first three stages of sleep refer to N1, N2, and N3 sleep; all are Non-Rapid eye movement (NREM) sleep. They account for 84% of adult sleeping time. Rapid eye movement (REM) sleep is the fourth stage and accounts for only 18% of sleeping time. During NREM sleep, many physiological changes take place. They start with the N1 stage, the lightest sleep stage, and account for 18% of sleeping time. N2 stage accounts for about 48% of the sleeping time; it starts when sleep begins to deepen. The deepest stage of sleep (N3) starts when brain waves are slow, and as a result, the body temperature drops down, and the heart rate slows. During NREM sleep, the brain activity will be similar to the awake state; therefore, it's called paradoxical sleep. This last stage is marked by increasing heart rate, blood pressure, and muscle atonia (4). With ageing, these sleeping stages are negatively impacted, which leads to disturbed sleep. There is also a natural decrease in total sleep time throughout the individual's life. Nighttime sleep rapidly decreases during early lifetime from 10 to 14 hours in the paediatric population to 6.5 to 8.5 in young adults. Nighttime sleep continues to decline at a slower rate to almost 5 to 7 hours a night. The natural decrease in nighttime sleep may lead to irrational expectations, producing more stress and anxiety that may negatively affect sleep or even cause insomnia (4)(5).

Insomnia is often described as difficulty in initiating sleep, with frequent nighttime or early morning waking and inability or difficulty returning to sleep. According to the (American Academy of Sleep Medicine), insomnia is diagnosed when the above symptoms present three nights per week for at least three months (6). Other more widely used criteria to diagnose insomnia are defined in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). According to DSM-IV, the requirements of primary insomnia are, firstly, for at least one month, difficulty initiating or maintaining sleep or nonrestorative sleep. Secondly, there is a significant impairment of important functioning areas. Thirdly, the disturbance does not occur exclusively during narcolepsy, breathing-related sleep disorder, circadian rhythm sleep disorder, or parasomnia. And lastly, it may not occur solely during another mental illness or a direct effect of substances use (1)(7).

Insomnia is the commonest sleep disorder in elderly. About 50% of older adults experience difficulty in initiating or maintaining sleep. The global prevalence of insomnia-related symptoms in old adults ranges from 38% to 48%, while the prevalence of insomnia disorder ranges from 12% to 20% (4). Insomnia poses an essential threat to healthy ageing. Research showed that elderly with untreated insomnia have a 23% increased risk of developing mental illness (4)(5). Moreover, insomnia is also associated with increased social and healthcare costs (1). Consequently, there is an increasing need to develop and validate methods to detect insomnia.

The diagnosis of insomnia is primarily made based on clinical information. However, objective methods like polysomnography and wrist actigraphy, and subjective methods like questionnaires have also been developed (4)(8). Polysomnography is used mainly to evaluate sleep apnea and parasomnias and is not recommended for the assessment of primary insomnia. Wrist actigraphy is used mainly to measure the response to treatment (8). A limited number of questionnaires have been validated to be used in the elderly (4). In 2016-2017 systematic review was performed to investigate the psychometric properties of sleeping questionnaires in elderly. The study concluded

that a relatively limited number of sleeping questionnaires were investigated in elderly, with strongly variable psychometric properties. Therefore, additional investigations were necessary for this specific group of population. This review tried to review and update the valid and reliable questionnaires to use as a diagnostic tool in elderly.

Methods

Search strategy

We used the search strategy developed in the paper by Degryse et al (9). The PRISMA guidelines were used to support the search strategy. A 'PIRT' (Patient - Index test - Reference test - Target condition), was created. The systematic literature search was in multiple databases, including PubMed, CINAHL, and Embase, from November 2016 to February 2017. We entered the same search terms in PubMed for the current review, from March 2017 to June 2021.

Inclusion and exclusion criteria

This review investigated the sleeping questionnaires used to judge or evaluate sleep in the elderly. We included studies that were conducted on patients aged 65 and higher. Because many studies did not define clear age limits, we included articles when the mean age of the studied population was above 60. We did not discriminate between studies conducted on community-dwelling, institutionalised, or hospitalised elderly. This study reviewed questionnaires judging the sleep quality; therefore, we excluded all studies investigating daytime sleepiness, underlying causes of insomnia, or sleeping-related disorders. We included studies that tried to validate a questionnaire or investigate its reliability. Both internal or external validation studies were included.

Selection studies

All studies collected according to the search strategy were reviewed and selected according to exclusion and inclusion criteria by title and abstract. The selection was carried out independently by two researchers. Articles that could not be selected based on title and abstract or where there was doubt about the inclusion and exclusion criteria were analysed according to the full article independently by both researchers. In case of persistent doubt, the articles were not included in the review.

Quality

We applied Quadas-2 to assess the quality of the included studies (10). We assessed the articles on risk of bias and applicability by answering specific questions regarding each particular domain. Those domains include patient selection, index test, reference standard, and flow of patients through the study.

Results

Search results

Our search from March 2017 to 30 June 2021 identified 2.853 articles in PubMed. All these articles were reviewed based on title and abstract by two independent researchers. 2.825 articles were excluded based on inclusion and exclusion criteria. The remaining 28 articles were analysed by full text. 4 articles met our inclusion criteria. The search flow is listed in figure 1. Quality of these articles was assessed by using the QUADAS checklist. In figure 2 the tabular presentation for QUADAS-2 results is mentioned. The 4 remaining articles studied psychometric properties of Athens Insomnia Scale(11), Insomnia Severity Index(11), PROMIS Sleep Disturbance Scale(12), Richards-Campbell Sleep Questionnaire (RCSQ-C)(13) and Pittsburgh Sleep Quality Index(14). Sleep questionnaires of previous review (9) were included to summarise questionnaires for use as a diagnostic tool for elderly.

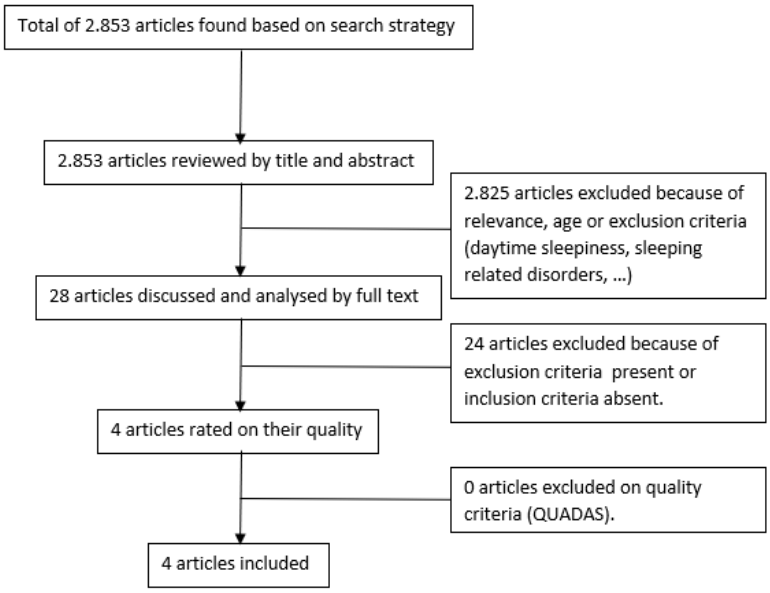


Figure 2: Selection Flowchart

Study	Risk of bias				Applicability concern		
	Patient selection	Index test	Ref standard	Flow & timing	Patient selection	Index test	Ref standard
(Lin *et al.*, 2020)	↓	↓	↑	↑	↓	↓	↑
(Full *et al.*, 2019)	↓	?	?	?	↓	↓	?
(Chehri *et al.*, 2020)	↓	↑	↓	↑	↓	↓	↓
(Chen *et al.*, 2019)	↓	?	↓	↑	↓	↓	↓

↓=Low risk, ↑=High risk, ?= Unclear risk

Figure 2: QUADAS-2

Study characteristics

The four new studies that were included described reliability and validation of 5 different sleep questionnaires. The following questionnaires were described; Athens Insomnia Scale (11), Insomnia Severity Index (11), PROMIS Sleep Disturbance Scale (12), Richards-Campbell Sleep Questionnaire (RCSQ-C) (13) and Pittsburgh Sleep Quality Index (14). Two studies (11)(13) only included ill patients. Both included only patients hospitalised at university hospitals. One of them only included critically ill patients in ICU (13). One study included elderly living independently or in continuing care retirement communities (12). The fourth study selected participants randomly with age 60 year and older (14). All 4 studies sought to analyse internal consistency, reliability and validation of the used sleep questionnaires.

The study of Chung-Ying Lin, 2019 (11) used different scales (Edmonton Symptom Assessment Scale, Hospital Anxiety and Depression Scale, Epworth Sleepiness Scale, the Pittsburgh Sleep Quality Index) and Actigraphy to validate the ISI and AIS. Different correlations with external data gathered from these scales were conducted to understand associations between the AIS and the ISI. Actigraph data were used to evaluate cut-offs for ISI and AIS. More specifically the impact of taking opioid or non-opioid analgesics was measured.

The other study (13), which only included ill patients more specifically in intensive care units, aimed to evaluate the reliability and validity of the Chinese version of the Richards-Campbell Sleep Questionnaire (RCSQ-C). The St Mary's Hospital sleep questionnaire, a scale designed to assess an inpatient's previous night's sleep, was used as convergent validity to measure the construct validity. An expert panel judged the content validity of the RCSQ-C.

The PROMIS Sleep Disturbance Scale (12) was assessed for elderly, with an average age of 83.6 years, who were recruited from retirement communities. The participants were mostly women 72.3%. Theoretically based correlations such as depression, stress, pain, and reduced QoL were included in the analyses to determine the reliability and validity of the sleep questionnaires. Various scales were used to gather these data: 'Perceived Quality of Life Scale (PQOL)', 'Center for Epidemiological Studies Depression Scale (CES D-10)', Cohen's Global Measure of Perceived Stress Scale' and 'PROMIS 6-item Pain Interference Scale'.

The fourth (14)and most recent study evaluated the validity and the reliability of the Persian version of the Pittsburgh Sleep Quality Index. In total 598 elderly people were selected to participate.

Athens Insomnia Scale

Two studies (11)(15) investigated the AIS scale. In the study of the previous review(9) the results of the subjective AIS scale were compared to actigraphy. Actigraphy was valued as objective sleep measurement. The study observed discrepancies between the subjective sleep questionnaire and actigraphy. These discrepancies were described in earlier studies. Similar discrepancies were detected in the studied population of patients with early- and moderate-stage Alzheimer dementia (15). Results of the study suggest that these discrepancies between AIS and actigraphy were not related to cognitive functioning as measured with MMSE. Using AIS as a screening tool for objective sleep disturbances measured by actigraphy was not recommended as the predictive value of the questionnaire is marginal for the control group of healthy elderly people and virtually nonexistent in the studied group of patients with early- and moderate-stage Alzheimer dementia. Reliability of the scale was not examined. Authors of the study concluded that patients with sleep problems may remain undetected and untreated using the questionnaire. The second study that investigated AIS (11), published in 2019, compared actigraph data, golden standard, with results of the AI scale of patients with advanced cancer. Internal consistency reliability was examined. Construct validity was investigated by conducting a confirmatory factor analysis. A score of 7/24 was set as the best cut-off score for AIS with a test–retest reliability using ICC of 0.80. The calculated sensitivity was 86% and specificity was 81%. The AUC for the AIS was 0.86 (95% CI = 0.82–0.90).

Insomnia Severity Index

In the previous review 2 studies (16)(17) evaluated psychometric properties of Insomnia Severity Index. The study of 2005 (16) examined in cancer patients the scale in its ability to screen for patients diagnosed with insomnia based on the Insomnia Interview Schedule. Reliability showed an overall Cronbach's alpha of 0.90. The study showed a score of 8 as the best cut-off score. With cut-off 8 the calculated sensitivity was 94.7% and specificity was 47.4%. The AUC obtained was 0.86 (95% CI = 0.81–0.91). The Chinese study of Yu et al. (17) with Cronbach's alpha of 0.81 for Chinese community-dwelling older people indicated high internal consistency. Construct validity was examined by hypothesis testing. Moderate positive relationship with CPSQI and moderate negative relationship with sleep efficiency was found. Discriminant validity was supported. No sensitivity or specificity was calculated. More recent study (11) published in 2019 with a patient group with advanced cancer suggested a cut-off of 9 with 86% sensitivity and 83% specificity. The calculated AUC was 0.82 (95% CI = 0.78–0.87).

Author, year	Country	Type of Questionnare	Study population	Age (mean +/- SD)	Sample size	Amount of elderly	Type of validation	Reliability and mean +/- SD scored on questionnaire	Standard error of measurement SEM (+/- 95% CI)	Sensetivity & specificity	AUC 95% CI
Chung-Ying Lin, 2019	Iran	Athens Insomnia Scale Insomnia Severity Index	patients with histologically verified malignant disease	61.3 +/-10.7	N=573	/	Reliability: internal consistency, composite reliability, test-retest reliability Validity: Unidimensional	AIS: α: 0.82 9.83 +/-4.10 ISI: α: 0.79 α: 10.54 +/-3.48	AIS: 2.57 On a total score of 24 ISI:2.00 On a total score of 28	AIS: Sens: 86% Spec: 81% ISI: Sen: 86% Spes: 83%	AIS: 0.86 0.82–0.90 ISI: 0.82 0.78–0.87
Savard M, 2004.	Canda	Insomnia Severity Index	Patients with cancer 24-90j	63.2 +/- 10.5	N= 1670	N=691 (41.4%)	Reliability: Internal consistency, Test - retest. Validity: Construct, Content. Polysomnography	α: 0.90 7.3 +/- 6.3	1.99 (+/- 3.90) On a total score of 28.	Sens: 94.7% Spes: 47.4	/
Yu D et al.	China	Insomnia Severity Index	Community dwelling elderly 60+	74.3 +/-7.3	N= 585	N=526 (+/-90%)	Reliability: Internal consistency. Validity: Construct, Content.	α: 0.81 10.38 +/- 5.23	2.28 (+/- 4.47) On a total score of 28.	/	/

α = Cronbach's alfa. SD = Standard deviation. CI= Confidence interval. PSI = Person Separation Index. AUC Area under the curve . SD: Standard deviation. CI: Confidence interval

Table 1: Insomnia Severity Index: Study characteristics

PROMIS Sleep Disturbance Scale

Only one study (12) was found that assessed psychometric properties of the PROMIS Sleep Disturbance Scale. The previous review did not include any studies with this scale. Reliability with a Cronbach's α of 0.86 illustrated a high internal consistency. Construct validity was analysed using exploratory factor analysis techniques and showed strong validity.

Richards-Campbell Sleep Questionnaire (RCSQ-C)

The RCSQ, a questionnaire used to assess sleep for hospitalised patients (13), showed a high internal consistency with Cronbach's α of 0.92. Content validity was established by an expert panel. Correlation coefficients supported the construct validity. Added by discriminant validity analysis with mean score of 49.34±24.64.

Pittsburgh Sleep Quality Index

6 studies included in the previous review analysed psychometric properties of the PSQI scale. By three of them (18)(19)(20) internal consistency was assessed with Cronbach's a ranging from 0.69-0.81. The study of Spira et al.(20) yielded adequate internal consistency with Cronbach's α of 0.69. Construct validity was validated by modest to moderate correlations with data of actigraphy and relevant variables such as sleep quality ($r=0.57$, $p<0.001$), latency ($r=0.45$, $p<0.001$), duration ($r=0.36$, $p<0.001$), efficiency ($r=0.46$, $p<0.001$), disturbances($r=0.36$, $p<0.001$), medications ($r=0.28$, $p<0.001$)

and daytime dysfunction (r=0.25, p<0.001). Correlations between the questionnaire and actigraphic were modest but statistically significant (daytime napping r=0.07 (p<0.001), wake after sleep onset r=0.18 (p<0.001)). The study of Most et al. (15), which included 45 elderly people, concluded that PSQI was not a good screening tool for insomnia. Regression analyses revealed no predictive value of subjective sleep parameters for objective sleep estimates by actigraphy. The study of Grandner et al. (21) shared the same conclusion. The PSQI global score didn't demonstrate a significant correlation with objective actigraphic measures (latency, total sleep time, wake-time-after-sleep onset or sleep efficiency). Internal homogeneity of the PSQI global score was evaluated with significant correlation (r=0.372 - r=0.804) of all components except one. And subjective sleep diary data are significantly correlated with PSQI score. The study of 2006 (21) validated the factor structure of the index. A 3-factor model is statistically favored over a single score. More recent of Beaudreau et al. (19), published in 2011, showed adequate internal consistency with Cronbach's α of 0.78. Construct validity was demonstrated by significant correlations with the Epworth Sleepiness Scale (r=0.11, p<0.001) and theoretically related constructs such as IADL scale (r=0.18, p<0.001), GDS (r=0.31, p<0.001) and objective actigraphy measurements (daytime inactivity (r=0.05, p=0.01) and wake after sleep onset (r=0.14, p<0.001)). . Although correlations with objective sleep measured by actigraphy were significant they were rather small. This study only included women ageing above 70. The German study (18) that evaluated psychometric properties of the PSQI didn't have separate public data of elderly participants. Psychometric properties for the study group couldn't be evaluated. Most recent study (14) confirmed the three-factor structure of PSQI. The study yielded a high internal consistency with Cronbach's α of 0.81. The index has significant correlations with other questionnaires (total score of insomnia severity index (r=0.625, p=0.001), Epworth sleepiness index (r=0.139, p=0.001), Berlin Index (r=0.336, p=0.001), and the Global Sleep Assessment Index (r=0.634, p=0.001)) which demonstrated its validity.

Author, year	Country	Type of Questionnare	Study population	Age (mean +/- SD)	Sample size	Amount of elderly	Type of validation	Reliability and mean +/- SD scored on questionnaire	Standard error of measurement SEM (+/- 95% CI)	Senseti vity & specifici ty	AUC
Azita Chehri, 2020	Iran	PSQI	Elderly men 60+	68.33 +/- 8.75	N= 598	N=598 (100%)	Reliability: Internal consistency	/	7.69+/-4.06	/	/
Beaudreau S, 2012.	USA	PSQI	Study of the osteoporotic fractures women 70-90+	83.4 +/-3.7	N= 2968	N= 2968 (100%)	Reliability: Internal consistency. Validity: Construct actigraphy	α: 0.78 6.3 +/-3.6	1.69 (+/- 3.31) on a total score of 21	/	/
Cole J, 2005.	USA	3 factor scoring model for the PSQI	Community-dwelling depressed and non-depressed 60-95j	68.90 +/- 6.34	N= 417	N≈304 (+/-73%)	Validity: construct	/	/	/	/
Grandner M, 2006.	USA	PSQI	Diverse 59-75j	66 +/- 4.8	Divers N= 53	N≈30 (+/- 58%)	Validity: Criteria Actigraphy	/	/	Sens: 89.6% Spec: 86.5%	/
Hinz A, 2017.	Germany	PSQI	Population study 18-80j	56.3 +/- 12.4	N= 9284	N=2621 (28.2%)	Reliability: internal consistency. Validity: construct	α: 0.75 5.14 +/- 3.44	1.72 (+/- 3.37) on a total score of 21	/	/
Most E, 2011	The Netherlands	PSQI, Sleep Disorder Questionnair, Athens Insomnia Scale	Alzheimer disease Controls	70.4 +/- 3.2 73.0 +/-4	N= 55 N = 46	N≈ 52 (+/-95%) N≈45 (+/-98%)	Validity: actigraphy	/	/	/	/
Spira A, 2011.	USA	PSQI	Elderly men 65+	76.4 +/- 5.5	N= 3059	N= 3056 (100%)	Reliability: Internal consistency. Validity: Construct actigraphy	α: 0.69 5.6 +/- 3.2	1.78 (+/- 3.49) on a total score of 21	/	/

α = Cronbach's alfa. SD = Standard deviation. CI= Confidence interval. PSI = Person Separation Index. AUC Area under the curve . SD: Standard deviation. CI: Confidence interval

Table 2: Pittsburgh Sleep Quality Index PSQI: Study characteristics

Minimal Insomnia Symptom Scale

The study of Westergren et al. (22) showed a reliability of 0.59, lower as typically considered desirable (0.70). Most optimal cut-off MISS score suggested ≥6 for elderly with a prevalence of 30.2%. More recent study of Hellstrom et al. (23) assessed the Minimal Insomnia Symptom Scale as a screening tool for elderly. Reliability was 0.81. A ROC-curve is calculated with area under the curve of 0.935 (95% CI: 0.90-0.97; p<0.001). As most optimal a cut-off MISS score of ≥7 was calculated with sensitivity of 93% and specificity of 84%. Three criteria of the Uppsala Sleep Inventory (USI), daytime sleepiness, feeling physically tired after sleep and experiencing sleep difficulties, were used to create a variable as a proxy gold standard.

Author, year	Country	Type of Questionnare	Study population	Age (mean +/- SD)	Sample size	Amount of elderly	Type of validation	Reliability and mean +/- SD scored on questionnaire	Standard error of measurement SEM (+/- 95% CI)	Sensetivity & specificity	AUC 95% CI
Hellström A, 2010.	Sweden	Minimal Insomnia Symptom Scale (MISS)	Diverse 65-90+	77.8 +/- 8.63	N= 548	N= 548 (100%)	Reliability: Internal consistency Validity: Criterion.	α: 0.81 4.1 +/- 2.8	1.22 (+/- 2.39) On a total score of 12	/	/
Westergren A, 2015.	Sweden	Minimal Insomnia Symptom Scale (MISS)	Diverse 65-90+	77.8 +/- 8.63	N= 548	N= 548 (100%)	Reliability: Item Response Theory, Internal consistency	PSI: 0.59 4.1 +/- 2.8	1.79 (+/- 3.51) On a total score of 12	/	/

α = Cronbach's alfa. SD = Standard deviation. CI= Confidence interval. PSI = Person Separation Index. AUC Area under the curve . SD: Standard deviation. CI: Confidence interval

Table 3: Minimal Insomnia Symptom Scale (MISS): Study characteristics

SCOPA-NS scale

The reliability in the study of Marinus et al. (24)was calculated with a Cronbach α of 0.88. Validity was assessed by comparing with other questionnaires such as the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Scale. Significant correlations were found with PSQI total score (r=0.83, p<0.001), with global sleep quality score (r=0.85, p<0.001) and with Epworth Sleepiness Scale (r=0.81, p< 0.001). The article considered a best cutoff value of 6/7 with an area under the ROC curve of 0.90, a sensitivity of 97% and a specificity of 80%. The receiver operating characteristic of this cutoff value was determined by using PSQI cutoff 8/9 as an external criterion. Another article (25), published in 2008, who studied psychometric properties of the SCOPA-NS scale determined Cronbach's alpha of 0.83. Proving a high internal consistency. Best cutoff was considered 6/7 with AUC of 0.96, sensitivity of 97% and specificity of 88%.

Author, year	Country	Type of Questionnare	Study population	Age (mean +/- SD)	Sample size	Amount of elderly	Type of validation	Reliability and mean +/- SD scored on questionnaire	Standard error of measurement SEM (+/- 95% CI)	Sensetivity & specificity	AUC 95% CI
Marinus J, 2003.	The Netherlands	SCOPA-NS scale	Parkinson's disease Controls	65.6 +/- 10.8 61.4 +/- 11.2	N=143 N=104	N=78 (66%) N=40 (41%)	Reliability: Internal consistency, Test-retest. Validity: Construct.	α: 0.88 4.9 +/- 4 2.8 +/- 2.7	1.50(+/- 2.94) On a total score of 15 0.94(+/- 1.84) On a total score of 15	Sens: 97% Spec: 80%	0.94
Martinez-Martin P, 2008.	Spain	SCOPA-NS scale	Parkinson's disease 40+	64.1 +/- 10	N= 187	N=78 (41.71%)	Reliability: Internal consistency. Validity: Construct.	α: 0.83 4.34 +/- 3.44	1.42 (+/- 2.78) On a total score of 15	Sens: 97% Spec: 88%	0.96 0.93–0.98

α = Cronbach's alfa. SD = Standard deviation. CI= Confidence interval. PSI = Person Separation Index. AUC Area under the curve . SD: Standard deviation. CI: Confidence interval

Table 4: SCOPA Nighttime Sleep Scale: Study characteristics

Women's Health Initiative Insomnia Rating Scale

Internal consistency of the Women's Health Initiative Insomnia Rating Scale (26) demonstrated an adequate with a Cronbach α of 0.78. Construct validity was explored by correlations with related constructs, Center for Epidemiologic Studies Depression Scale (CES-D), the RAND-36 subscales, the climacteric symptoms, and the Negative Emotional Expressiveness Scale (NEE). Although the correlations were in the hypothesised direction none of the reported correlations are large.

Comprehensive Sleep Questionnaire

Internal consistency was proven very high with Cronbach's α of 0.94. Construct validity was analysed by calculating correlations with clinical variables. Statistically significant correlations were found for Mini-Mental State Examination (MME) (P=0.04), Activities of Daily Living (ADL) (p=0.01) and Neuropsychiatric Inventory and Clinical Dementia Rating (CDR) (p=0.001).

Insomnia in the Elderly Scale

The study (27) used the diagnostic criteria for insomnia given in the DSM-IV-TR and the ICD-10 as the golden standard. Psychiatric interviews with all participants reviewed this criterion to assess insomnia in the study population. Reliability analysis of the scale showed a correlation coefficient of 0.93 (95% CI 0.859–0.964). Most suitable cutoff to detect insomnia was 3/4 with AUC of 0.87 (95% CI 0.843–0.893), sensitivity of 86% and specificity of 70%.

Medical Outcomes Study Sleep Scale

The study (28) was divided in two subscales, Sleep Problems Index I and Sleep Problems Index II. adequate to high internal consistency was demonstrated with Cronbach α values of 0.72 and 0.82. Concurrent validity was assessed by statistically significant correlations with the PSQI (r ranging from 0.33 to 0.74, P=0.001). Index I with a cutoff point above 22.22 had a sensitivity of 80% and a specificity of 74%. For Index I a score above cutoff 33.33 had a 60% sensitivity and a 91% specificity.

Author, year	Country	Type of Questionnare	Study population	Age (mean +/- SD)	Sample size	Amount of elderly	Type of validation	Reliability and mean +/- SD scored on questionnaire	Standard error of measurement SEM (+/- 95% CI)	Sensetivity & specificity	AUC 95% CI
Kelsie M. Full, 2017	USA	PROMIS Sleep Disturbance Scale	Older adults in independent-living & continuing care retirement communities	83.6 +/-6.4	N=307	N=307 (100%)	Reliability: internal consistency reliability	α: 0.856 49.60 +/- 7.91	/	/	/
Li-xia Chen, 2018	China	Richards-Campbell Sleep Questionnaire (RCSQ-C)	Critically ill patients in university teaching hospital ICUs	64.74 +/- 16.15	N=150	/	Internal consistency reliability, Inter-rater reliability/agreement	α: 0.923 49.34 +/- 24.64.	/	/	/
Levine D, 2003.	USA	Women's Health Initiative Insomnia Rating Scale	Postmenopauzal women 50-79	Group 1: 62.07 +/- 7.41 Group 2: 67.7	N= 66 269 N= 459	Groep1: N= 29 821 (45%) Groep 2 N=	Reliability: Internal consistency, Test-retest. Validity: Construct. Reliability: Internal consistency. Validity: Actigraphy	Group 1: α: 0.786 6.61 +/- 4.45 Group 2: α: 0.794 7.26 +/- 4.92	Group 1: 2.06 (+/- 4.04) On a total score of 20 Group 2: 2.23 (+/- 4.37) On a total score of 20	/	/
Manni R, 2013.	Italy	Comprehensive Sleep Questionnaire	Alzheimer's disease	77 +/-5.8	N= 140	N= 137 (+/- 98%)	Reliability: Internal consistency. Validity: Construct.	α: 0.938 16.23 +/- 4.9	1.22 (+/- 2.39) On a total score of 28	/	/

α = Cronbach's alfa. SD = Standard deviation. CI= Confidence interval. PSI = Person Separation Index. AUC Area under the curve . SD: Standard deviation. CI: Confidence interval

Table 5: Other scales: Study characteristics

Discussion

Many sleeping questionnaires are used to define sleep quality and facilitate early detection of sleeping disorders in elderly. Many sleeping questionnaires are used for this purpose therefore it's crucial to define the validity and the reliability of these scales to be used in that specific group of population.

Reliability

To evaluate the reliability of an instrument we looked at the value of Cronbach's alpha, which is a reliability coefficient based on internal consistency. 0.70 is the threshold over which a questionnaire can be considered reliable. the homogeneity of the items increases with increasing Cronbach's alpha. (29). The SEM value also provides good information about reliability. A smaller SEM value indicates better reliability making it easier to compare the reliability of multiple scales. Starting with the PSQI, three studies (18)(19)(20) showed an adequate internal consistency, with a good Cronbach's alphas. However, most of the PSQI subscales showed a good correlation with the global score except for the Medications and Daytime Dysfunction subscales suggesting excluding these two subscales when PSQI is used to assess sleep quality in this specific group of population. Regarding ISI, included studies showed acceptable internal consistency with a good Cronbach's alphas. However, two studies (16)(17) investigated ISI in cancer patients. One (16) of the mentioned two studies was conducted in a hospital setting. The results suggested that the ISI is a reliable instrument for evaluating the severity of insomnia secondary to cancer in hospital and general clinic settings. The SCOPA night scale encounters more than adequate internal consistency but only applicable for sleep disorder in Parkinson's disease. One (22) of two studies (22)(23)that assessed internal consistency of the Minimal Insomnia Symptom Scale (MISS) showed an inadequate reliability with Cronbach's alpha of 0.59. Authors suggest that this is due to the low number of items in the scale. In contrast Cronbach's alpha of other study (23) is 0.81 which displays an adequate internal consistency. More research is necessary.

Validity

On the other hand, validity is a more complex concept. Basically, validity is divided into 3 main types: content, criterion and construct validity. Unfortunately, there is no gold standard to determine criterion validity of sleep quality. Some articles suggest actigraphy or psychiatric interviews to DSM diagnostic criteria as the golden standard for insomnia. Mostly construct validity was validated with correlations of relevant theoretically variables found in the literature. Authors measured if the scales behave consistent with these variables who hypothetically represent an insomnia diagnosis. These construct variables can differ for each study depending on the choice of the authors and what they find in the literature. This makes comparison of the strength of the validation of different questionnaires very difficult. For example, the PSQI scale is validated mostly with significant correlations of theoretically relevant variables. The hypotheses of the construct were based on variables such as sleep diary data, depression scales, quality of life questionnaire, physical complaints, anxiety, fatigue, etc. Moderate to high correlations were determined for these theoretical variables. However even if criterion validity shows significant correlations with actigraphy for some measures they show only poor correlations in all three articles (15)(18)(20) who compared the scale with actigraphy data. Not only with PSQI can we see that there is a difference in the validation of the construct by objective or subjective measurements of sleep. Some correlations with actigraphy measures were significant but at most moderate for the Insomnia Severity Index. Overall objective measurements as construct, if significant, tend to show less correlation than subjective measurements. As there is no generally acceptable gold standard for diagnosing sleep quality the question remains how we can define sleep quality at all? Can objective measurements only determine sleep quality? Is sleep quality not a perception that is subjective and so difficult to measure? Interesting is to note that most constructs included variables of physical functioning (e.g.

Mobility/IADL, physical complaints), cognitive functioning (e.g. depression, MMSE) and some social functioning. Construct validity of several questionnaires was often validated by correlation with other validated sleep questionnaires.

Sensitivity and specificity

If we want to use any questionnaire as a diagnostic test, it is crucial to determine its sensitivity and specificity. As mentioned above, because there is no generally accepted gold standard, studies used different external criteria for calculating the AUC. This discrepancy in calculation of AUC makes comparaison in questionnaires challenging. The sensitivity and specificity is crucial to decide the best screening tool for insomnia by elderly. However, a very limited number of studies had mentioned the specificity and the sensitivity of the investigated questionnaire. Starting with PSQI from the six included studies in the previous review(15)(19)(30)(21)(15)(20) and 1 study (14) in this review no article had determined the sensitivity and specificity for the studies group. Even though PSQI is seen as the most popular screening tool in the general population, data for use in the elderly population isn't available. For patients with malignancy ISI sleeping questionnaires showed promising sensitivity in both hospitalised and general settings. ISI and AIS, both showed a good sensitivity of 86% and specificity of 81% and 83% respectively in hospitalised patients with malignancy. However, a community-dwelling study in 2004 (16) applied to patients with malignancy mentioned a high sensitivity of 95% and a low specificity of 48%. Only 41% of the studied population was above the age of 65 years. A more recent study (11)mentioned specificity of 86% and sensitivity of 81%. The percentage of the elderly was not mentioned in the published article. Moreover, both studies have been applied to patients diagnosed with malignancy. SCOPA NS have shown the highest sensitivity 97% among the other questionnaires with 80% specificity. Two studies (24)(25) mentioned the same sensitivity and specificity, one limitation is that both studies have been applied to patients with Parkinson's disease. One study (11) calculated the sensitivity and the specificity of AIS 86% and 81% respectively. There was a remarkable heterogeneity in the studied population. Therefore, it was difficult to generalise the specificity and the sensitivity of the studied questionnaires. The validation of the Insomnia in the elderly scale used the psychiatric interview (DSM-IV-TR diagnostic criterion A) as the gold standard. A sensitivity of 86.4% and a specificity of 69.5% was determined for a population of 926 non-institutionalized elderly aged at least 65years. By using an acceptable golden standard in a large homogeneous population of elderly this questionnaire can be used as a screening tool for the studied population. The evaluation (23) of the Minimal Insomnia Symptom Scale published in 2010 has a better calculated sensitivity of 93% and specificity of 84%. A study of 548 elderly individuals representing the ageing population of Sweden. Limitation of this study was the use of a proxy golden standard to calculate the AUC. The Women's Health Initiative Insomnia Rating Scale (WHIIRS) has worse sensitivity and specificity. The studied population (26) of only (postmenopausal) women makes scale not widely applicable for all elderly population. Also, clinical diagnosis of insomnia as the gold standard wasn't adopted. Another limitation (26) was that the study only included postmenopausal women with low percentage of elderly (+/-24 elderly) therefore not possible to generalise of all elderly population.

Conclusion

Not a lot of research is available for the general practitioner to determine a good insomnia screening tool adapted for the elderly population. Although many sleep questionnaires widely used in the general population are validated and have adequate internal consistency for use in the studied population, exploration as screening tool is minimal. The Insomnia in the Elderly Scale and the Minimal Insomnia Symptom Scale (MISS) display promising results but lack reproducibility. MISS had an inadequate Cronbach's α provoking doubts for its internal consistency. An important limitation to comparing questionnaires, is the absence of a general accepted gold standard for insomnia. Discrepancy in subjective and objective measurements to determine insomnia.

Bibliography

1. Roth T. Insomnia: Definition, Prevalence, Etiology, and Consequences. 2007;3(5):3–6.

2. Da Silva AA, De Mello RGB, Schaan CW, Fuchs FD, Redline S, Fuchs SC. Sleep duration and mortality in the elderly: A Systematic review with meta-analysis. BMJ Open. 2016;6(2):1–13.

3. Dew MA, Hoch CC, Buysse DJ, Monk TH, Begley AE, Houck PR, et al. Healthy older adults' sleep predicts all-cause mortality at 4 to 19 years of follow-up. Psychosom Med. 2003;65(1):63–73.

4. Patel D, Steinberg J, Patel P. Insomnia in the Elderly : A Review. 2018;

5. Ohayon MM, Caulet M, Guilleminault C. How a general population perceives its sleep and how this relates to the complaint of insomnia. Sleep. 1997;20(9):715–23.

6. Jansson-Fröjmark M, Sunnhed R. Psychometric Properties of Two Brief Versions of Cognitive, Insomnia-Specific Measures: The Anxiety and Preoccupation About Sleep Questionnaire and the Sleep-Associated Monitoring Index. Psychol Rep. 2020;123(3):966–82.

7. Abuse S. IMPACT OF THE DSM-IV TO DSM-5 CHANGES ON THE NATIONAL SURVEY ON DRUG USE AND HEALTH. 2016;(June).

8. Marino M, Li Y, Rueschman MN, Winkelman JW, Ellenbogen JM, Solet JM, et al. Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep. 2013;36(11):1747–55.

9. Degryse PPJ. Psychometric properties of sleep questionnaires investigating sleep quality in older people : a systematic review. 2017;

10. Whiting PF, Weswood ME, Rutjes AWS, Reitsma JB, Bossuyt PNM, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol. 2006;6(October).

11. Lin CY, Cheng ASK, Nejati B, Imani V, Ulander M, Browall M, et al. A thorough psychometric comparison between Athens Insomnia Scale and Insomnia Severity Index among patients with advanced cancer. J Sleep Res. 2020;29(1):1–11.

12. Full KM, Malhotra A, Crist K, Moran K, Kerr J. Assessing psychometric properties of the PROMIS Sleep Disturbance Scale in older adults in independent-living and continuing care retirement communities. Sleep Heal [Internet]. 2019;5(1):18–22. Available from: https://doi.org/10.1016/j.sleh.2018.09.003

13. Chen L xia, Ji D hong, Zhang F, Li J hua, Cui L, Bai C jie, et al. Richards-Campbell sleep questionnaire: psychometric properties of Chinese critically ill patients. Nurs Crit Care. 2019;24(6):362–8.

14. Chehri A, Nourozi M, Eskandari S, Khazaie H, Hemati N, Jalali A. Validation of the

Persian version of the pittsburgh sleep quality index in elderly population. Sleep Sci. 2020;13(2):119–24.

15. Most EIS, Aboudan S, Scheltens P, Van Someren EJW. Discrepancy between subjective and objective sleep disturbances in early-and moderate-stage alzheimer disease. Am J Geriatr Psychiatry [Internet]. 2012;20(6):460–7. Available from: http://dx.doi.org/10.1097/JGP.0b013e318252e3ff

16. Savard, Marie-Hélène, Josée Savard, Sébastien Simard and HI. Empirical Validation of the Insomnia. Psychooncology. 2005;14(July 2004):429–41.

17. Yu DSF. Insomnia Severity Index: Psychometric properties with Chinese community-dwelling older people. J Adv Nurs. 2010;66(10):2350–9.

18. Hinz A, Glaesmer H, Brähler E, Löffler M, Engel C, Enzenbach C, et al. Sleep quality in the general population: psychometric properties of the Pittsburgh Sleep Quality Index, derived from a German community sample of 9284 people. Sleep Med. 2017;30:57–63.

19. Beaudreau SA, Spira AP, Stewart A, Kezirian EJ, Lui LY, Ensrud K, et al. Validation of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Scale in older black and white women. Sleep Med [Internet]. 2012;13(1):36–42. Available from: http://dx.doi.org/10.1016/j.sleep.2011.04.005

20. Spira AP, Beaudreau SA, Stone KL, Kezirian EJ, Lui LY, Redline S, et al. Reliability and validity of the pittsburgh sleep quality index and the epworth sleepiness scale in older men. Journals Gerontol - Ser A Biol Sci Med Sci. 2012;67 A(4):433–9.

21. Grandner MA, Kripke DF, Yoon IY, Youngstedt SD. Criterion validity of the Pittsburgh Sleep Quality Index: Investigation in a non-clinical sample. Sleep Biol Rhythms. 2006;4(2):129–36.

22. Westergren A, Broman JE, Hellström A, Fagerström C, Willman A, Hagell P. Measurement properties of the minimal insomnia symptom scale as an insomnia screening tool for adults and the elderly. Sleep Med [Internet]. 2015;16(3):379–84. Available from: http://dx.doi.org/10.1016/j.sleep.2014.10.016

23. Hellström A, Hagell P, Fagerström C, Willman A. Measurement properties of the Minimal Insomnia Symptom Scale (MISS) in an elderly population in Sweden. BMC Geriatr [Internet]. 2010;10(1):84. Available from: http://www.biomedcentral.com/1471-2318/10/84

24. Marinus J, Visser M, Van Hilten JJ, Lammers GJ, Stiggelbout AM. Assessment of sleep and sleepiness in parkinson disease. Sleep. 2003;26(8):1049–54.

25. Martinez-Martin P, Visser M, Rodriguez-Blazquez C, Marinus J, Chaudhuri KR, van Hilten JJ, et al. SCOPA-sleep and PDSS: Two scales for assessment of sleep disorder in Parkinson's disease. Mov Disord. 2008;23(12):1681–8.

26. Levine DW, Kripke DF, Kaplan RM, Lewis MA, Naughton MJ, Bowen DJ, et al. Reliability and validity of the Women's Health Initiative Insomnia Rating Scale. Psychol Assess.

2003;15(2):137–48.

27. Manni R, Sinforiani E, Zucchella C, Terzaghi M, Rezzani C. A sleep continuity scale in Alzheimer's disease: Validation and relationship with cognitive and functional deterioration. Neurol Sci. 2013;34(5):701–5.

28. Zagalaz-Anula N, Hita-Contreras F, Martínez-Amat A, Cruz-Díaz D, Lomas-Vega R. Psychometric properties of the medical outcomes study sleep scale in Spanish postmenopausal women. Menopause. 2017;24(7):824–31.

29. Thanthirige P, Shanaka R, Of A, Contributing F, Time TO, Of O, et al. No Title. 2016;334(August).

30. Cole JC, Motivala SJ, Buysse DJ, Oxman MN, Levin MJ, Irwin MR. Validation of a 3-factor scoring model for the Pittsburgh Sleep Quality Index in older adults. Sleep. 2006;29(1):112–6.

Persoonlijke reflectie paper

Een paper, het werk van lange adem. Onze paper is inderdaad niet op een drafje gemaakt. Initieel gestart met een totaal ander onderwerp, het nut van het gebruik van Elocom vij vaccinatie van kinderen. We hadden hiervoor literatuur onderzoek gedaan. Echter na de start van Corona leek het niet mogelijk om buisjes te verkegen om onze RCT op te starten. In die zelfde periode ben ik gestart bij K&G om ook daar de RCT op te starten. Hoewel we uiteindelijk niet konden verder gaan met het project ben ik hierdoor begonnen met het opnemen van zittingen bij K&G. Zittingen die ik tijdens mijn ziekenhuisstage en huidige HAIO stage steeds heb kunnen verderzetten. Kind en Gezin bied ik ook allerlei opleidingen aan die niet enkel een meerwaarde bieden tijdens de zittingen maar ook in de praktijk. Zo zijn het soms zeer kleine zaadjes die kunnen groeien tot iets moois. Hoewel het initiële onderwerp na een jaar stopgezet is ben ik zeer blij dat we zo gestart zijn.

Voordat het project van ons initiële onderwerp is stopgezet hebben we heel wat moeilijkheden gehad bij de communicatie met onze promotor. Als ik voor mezelf spreek heeft dit een serieuze knauw gegeven aan onze motivatie van de masterpaper. Halverwege in het HAIO schap zijn we halsoverkop gestart aan een paper met een volledig ander onderwerp. We hebben beiden niet de intentie om academische vervolgopleidingen te volgen en het schrijven van een masterpaper tijdens het Haio schap is voor mij dan ook geen prioriteit geweest. Er zijn dagen dat ik liever op het kabinet nableef omdat er thuis aan de paper moest geschreven worden. Voor mij geen geheim dat dit niet mijn ding is. Laat mij maar het klinisch werk doen in de huisartsenpraktijk!

Uiteindelijk gestart met het nieuwe onderwerp. Het actualisering van eerdere review begeleid door onze promotor. Zeer actuele thematiek in de huisartsenpraktijk vonden we allebei. Echter is niet veel later onze promotor op emiraat gegaan waardoor er voor ons nog teveel onduidelijkheden waren. Met onze nieuwe promotor hebben we opnieuw een doorstart kunnen maken. Het onderwerp was hem niet aantrekkelijk doch hebben we de begeleiding gehad om ons de motivatie te herwinnen. Taken hadden we goed verdeeld en Farkad en ik konden zo elk aan ons deel werken. We spraken regelmatig af om elkaar op de hoogte te houden waar we stonden dit ging goed. Daar is een zwangerschap tussen gekomen en drukke werkschema's van de praktijk.

Jammer dat we geen nieuwe zaken hebben ontdekt tijdens de actualiseren van voorgaande review. Een handige tool voor in de huisartsenpraktijk zit er nog niet in. Het zou een meerwaarde zijn geweest. Het is ook zeer moeilijk iets subjectiefs zoals slaap te vervatten in een af te vinken vragenlijst. Heel eigen aan de huisartsenpraktijk uiteraard.