

Discrepancies in self-reported and actigraphy-based sleep duration are associated with self-reported insomnia symptoms in community-dwelling older adults

Siobhan Scarlett, MSc^{a,*}, Hugh N. Nolan, PhD^a, Rose Anne Kenny, MD^{a,b}, Matthew D.L. O'Connell, PhD^{a,c}

^a The Irish Longitudinal Study on Ageing, Trinity College Dublin, University of Dublin, Ireland

^b Mercer's Institute for Successful Ageing, St James's Hospital, Dublin, Ireland

^c Department of Population Health Sciences, School of Population Health and Environmental Sciences, Kings College London, London, United Kingdom

ARTICLE INFO

Article History:
Received 8 January 2020
Revised 11 June 2020
Accepted 12 June 2020

Keywords:
Older population
Measurement agreement
Accelerometer
Insomnia
Sleep duration

ABSTRACT

Objectives: To establish agreement between self-reported and actigraphy-based total sleep time (TST). To determine the impact of self-reported sleep problems on these measurements.

Design: Cross-sectional study using data from Wave 3 of The Irish Longitudinal Study on Ageing (2014–2015).

Participants: Community-dwelling older adults, aged ≥ 50 years, with self-reported sleep information and ≥ 4 days of actigraphy-based TST ($n = 1520$).

Measurement: Self-reported total sleep time, daytime sleepiness, insomnia symptoms (trouble falling asleep, trouble waking too early) measured during a structured self-interview. Actigraphy-based total TST was collected using GENEactiv wrist-worn accelerometers. Demographic characteristics and health information were controlled for. Analyses included descriptive statistics, reliability and agreement analysis using paired *t*-tests, intra-class correlations and Bland-Altman analysis. Linear regression was used to model associations with measurement discrepancies.

Results: Participants reported that they slept 7.0 hours (SD: 1.4, Range: 2.0–13.0 hours) on average, compared to 7.7 hours (SD: 1.2 hours, Range: 3.0–13.0 hours) recorded by accelerometry. Trouble falling asleep or waking too early “most of the time” were associated with under-reporting of sleep by 2.3, and 2.2 hours respectively. Agreement between measurements had an intra-class correlation of 0.18 and wide 95% limits of agreement (-3.90 to 2.55 hours). Under-reporting of sleep was independently associated with insomnia symptoms.

Conclusion: The agreement between self-reported and actigraphy-based TST in community dwelling older adults was low. Self-reported insomnia symptoms were independently associated with under-reporting of sleep. Studies seeking to measure sleep duration should consider inclusion of questions measuring experience of insomnia symptoms to account for potential influence on measurements.

© 2020 The Authors. Published by Elsevier Inc. on behalf of National Sleep Foundation. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Introduction

Sleep is a restorative process and plays a vital role in preservation of cognitive, physical and mental health.^{1–3} Poor sleep quality and duration has been associated with adverse health outcomes such as cognitive impairment, cardiovascular disease, depression, and mortality, many of which older adults are at increased risk of.^{1,4–8} Measurement of sleep can be achieved both subjectively and objectively. Despite the importance of measurement of sleep quality and

duration, the number of studies examining agreement between types of measurement in large community samples is limited.^{9–14}

Surveys typically use subjective means such as sleep questionnaires, or sleep diaries, to assess sleep in participants. Laboratory based polysomnography is considered the gold standard for sleep measurement, but is expensive and impractical to consider for use in large community dwelling populations.¹⁵ Activity based monitoring through the use of accelerometer devices has been shown to be an effective tool for sleep research.¹⁶ These have become a feasible method for capturing objective sleep in large population studies. Wrist-worn devices are cost-effective and allow for non-invasive objective measurement of sleep in natural settings over long periods of time.

*Corresponding author: Siobhan Scarlett, MSc, The Irish Longitudinal Study on Ageing, Trinity College Dublin, University of Dublin, Trinity Central, 152–160 Pearse St, Dublin 2 D02 R590, Ireland.

E-mail address: sscarlet@tcd.ie (S. Scarlett).

Sleep quality is a complex construct. Perception of sleep may not corroborate with measured sleep, potentially impacted by sleep complaints experienced by an individual.^{9,17–19} A number of studies have identified disagreement between subjective and objective measurements of sleep in older adults where accelerometer devices were used.^{9–14} The CARDIA study showed subjective reports of sleep to be longer than actigraphy-measured sleep, increasing by 31 minutes for each additional hour of measured sleep.¹³

Sleep complaints are prevalent in the older population.²⁰ It has been reported that 30–48% of older adults report experiencing insomnia symptoms, with the most common complaint being difficulty initiating or staying asleep.²¹ Sleep complaints were found to drive differences between objective and subjective measurements. Using wrist worn accelerometer devices in a sample of adults aged 57–97 years and the Pittsburgh Sleep Quality Index (PSQI), van den Berg et al. showed that adults reporting poorer sleep were more likely to under estimate their sleep duration compared to their accelerometer measurement,¹¹ while McCrae et al. reported that subjective and objective measurements matched only for those without sleep complaints.¹⁴ Similarly, vulnerable older adults reporting sleep complaints were more likely to negatively report their sleep relative to their objective measurements.¹²

It is advised for studies to include both subjective and objective measurements of sleep.^{9,22} The inclusion of subjective measurement is common in population studies, but there is no consensus on which tool is optimal for this purpose. Subjective sleep duration is often captured by a single survey question such as “How many hours on average do you sleep per night”. A number of high quality sleep scales have been validated, such as the PSQI (19 items),²³ the Epworth Sleepiness Scale (8 items),²⁴ or the Munich Chronotype Questionnaire (13 items).²⁵ These scales provide detailed data on different aspects of sleep, but application may be difficult in studies with high participant burden or clinical settings with limited time.

Large studies examining agreement between subjective and objective measurements in community dwelling older adults are limited.^{11,26} A number of large, population representative studies of older adults have introduced accelerometry measurement including The Irish Longitudinal Study on Ageing (TILDA),²⁷ the English Longitudinal Study of Ageing,²⁸ and the Whitehall II study of British civil servants.²⁶ These studies provide unique opportunities to analyze sleep in rich, complex datasets. Using the Whitehall II cohort, van Hees et al. showed gender, depression and insomnia symptoms to be drivers of measurement disagreement.²⁶ As the use of objective measurement becomes more common in large population studies, it is important to understand the concordance of self-reported and objective sleep measurements as typically measured in these studies. The majority of research in this area has used small samples with limited generalisability.^{9,12,19,29}

This study set out to contribute to existing literature by assessing measurement agreement in TILDA cohort of community-dwelling older adults using wrist-worn accelerometry and quick to administer survey questions on sleep duration and complaints. The nature of the study allowed for an extensive list of known confounders for sleep measurement agreement to be accounted for in a large, population derived sample of community dwelling older adults.

This study aims to 1) establish the overall level of agreement between self-reported total sleep time and actigraphy-based total sleep time and 2) to determine the impact self-reported sleep problems have on measurement agreement using three short sleep problem questions.

Methods

Data were drawn from TILDA, which is a nationally representative population study of community dwelling adults aged 50 and over in

the Republic of Ireland. The TILDA study design has been outlined previously.^{27,30} Briefly, the sample was selected using the RANSAM sampling procedure.³¹ Interviewers visited residential addresses drawn from the Irish National Geodirectory. Any adult aged 50 or older, including their spouse of any age, were invited to participate. 8175 adults aged 50 or older took part in the first wave of data collection in TILDA (2009–2011) representing a response rate of 62%. Participants provided written, informed consent. Ethical approval was obtained from the Trinity College Dublin Faculty of Health Sciences Research Ethics Committee.

Through structured interviews, TILDA collects detailed information on the health, social and financial situation of each participant. Fig. 1 presents a detailed breakdown of the analysis sample for this study. This study analyzed data from Wave 3, collected during 2014 and 2015 ($n = 6902$). Where a participant could not complete the survey due to a cognitive or physical impairment, a proxy interview was offered. An end-of-life interview was administered to a close relative or friend where a participant had passed away. Only participants who were aged 50 years or older and had completed the self-interview were included ($n = 6497$). Wave 3 included a comprehensive health assessment.³² The health assessment was carried out by trained research nurses and offered to participants who completed the structured interview. 190 GENEactiv wrist-worn accelerometer devices were also available during Wave 3, facilitating an accelerometry study on a sub-sample of Wave 3 health assessment participants. Following their health assessment, a randomly selected group of participants were asked to wear an accelerometer device for seven consecutive days immediately following their assessment ($n = 1578$).³³

Previous analyses reported that four days of measurement were sufficient to obtain a reliable estimate of total sleep time in this cohort.³⁴ Due to technical faults with the device, or where a participant did not wear the device for the full seven days, 45 devices were returned with fewer than four days of data captured. A further 13 participants with incomplete data on self-reported sleep were excluded from the analysis. The final analysis sample consisted of 1520 participants.

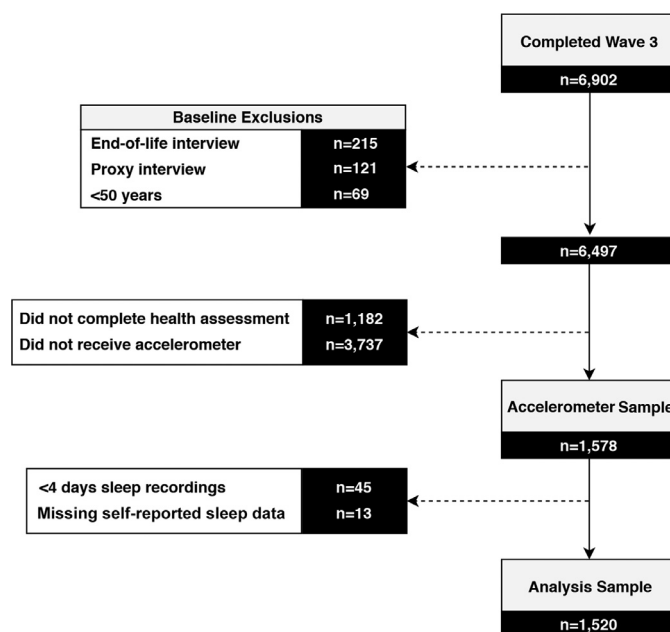


Fig. 1. Sample flowchart displaying exclusion criteria and breakdown of analysis sample.

Accelerometer devices

Devices used were wrist-worn GENEactiv devices. These devices have a measurement range of ± 8 g with a maximum logging period of 7 days at 100 Hz. The device is lightweight and water resistant. Devices have a body-temperature sensor that can confirm wear and nonwear periods. Data were processed using a fully automated Micro-Electro-Mechanical Systems (MEMS) accelerometer classification algorithm devised for epidemiological studies.^{34,35}

Measures

Total sleep time

Actigraphy-based Total Sleep Time (TSTobj) was classified as average total sleep time for the length of device wear as measured by the accelerometer device.^{34,35} Self-reported Total Sleep Time (TSTsub) was classified using the interview question “Approximately how many hours do you sleep on a weeknight?”. Participants were asked to round to the nearest hour when asked to self-report their average sleep. For comparative purposes, TSTobj was rounded to the nearest hour. As previous analyses have shown there to be no difference in weekday and weekend recordings in this sample, measured total sleep time was derived using all available days of recording.³⁴

Sleep reporting difference

A sleep reporting difference score was calculated as (TSTsub-TSTobj). A positive value indicated over-reporting of sleep relative to objective measurement, and a negative value indicated under-reporting.

Sleep problems

Daytime sleepiness

Participants rated whether they were likely to doze off or fall asleep during the day on a four point likert scale – 0.would never doze, 1.slight chance of dozing, 2.moderate chance of dozing, 3.high chance of dozing.

Insomnia symptoms

Participants were asked two additional questions drawn from the Jenkins Sleep Scale which relate to insomnia symptoms.³⁶ These asked how often the participant has trouble falling asleep, and how often they have trouble waking up too early and not being able to fall asleep before. These were answered on a three point likert scale – 0. rarely or never, 1.sometimes, 2. most of the time.

Covariates

Variables known to be related to sleep duration were included as potential confounders. Socio-demographic information including age, sex, education level (primary/none, secondary level, third level or higher), marital status (married, not married, separated/divorced, widowed), area of residence (urban, rural) and employment status (employed, retired/not employed) were collected during the interview. The season of recording for measured sleep was also included.

Cognitive status was assessed using the Mini-Mental State Examination (MMSE).³⁷

Health measures included self-rated health (excellent/very good/good, fair/poor), smoker status (smoker, non-smoker), self-reported pain and depressive symptoms as measured by scoring nine or higher on the eight-item center for Epidemiological Depression Scale.³⁸

Height and weight were measured during the health assessment. BMI was defined as weight divided by height squared (kg/m^2) and categorised as underweight/normal weight (0–24.9), overweight (25.0–29.9), and obese (≥ 30).

Participants were asked if they have ever been diagnosed with any chronic or cardiovascular conditions. Chronic conditions included lung disease, asthma, arthritis, osteoporosis, cancer, Parkinson disease, stomach ulcer, varicose ulcer, liver disease, thyroid disease, or kidney disease. Cardiovascular conditions included hypertension, stroke, angina, heart attack, heart murmur atrial fibrillation, or other abnormal heart rhythms. Participants were classified as having none, or one or more chronic conditions, and none, or one or more cardiovascular conditions.

Detailed medication use information was also collected. Medications were classified using Anatomical Therapeutic Chemical (ATC) classification codes. Use of medications that impact sleep were included as covariates. Antidepressant medication was classified by ATC code N06A. Use of sleep medication included ATC codes N05A (antipsychotics), N05B (anxiolytics), N05C (hypnotics and sedatives), and R06A (antihistamines). Use of antihypertensive medication was classified by ATC codes C02 (antiadrenergic agents), C03 (diuretics), C07 (β blockers), C08 (calcium-channel blockers), and C09 (angiotensin-converting enzyme inhibitors).

Statistical analysis

Analyses were performed using Stata version 15.1 (Stata Corp., College Station, TX). Descriptive statistics for sleep parameters and participant characteristics were produced. TSTsub and TSTobj differences are presented to examine where discrepancies were more pronounced within sample characteristics.

Intra-class correlation coefficients produced using two-way mixed-effects models and Bland-Altman analysis was used to examine measurement agreement between TSTsub and TSTobj.^{39,40} Overall agreement between measurements, and agreement by sleep problems were calculated.

The Bland-Altman analysis used TSTobj measurements in their original format prior to rounding. Linear regression models were used to analyze independent associations between TSTsub-TSTobj difference scores and sleep problems (daytime sleepiness, trouble falling asleep and trouble waking up too early). Basic models were adjusted for age and sex. Full models were additionally adjusted for covariates associated with large discrepancies in TSTsub and TSTobj. Full model outputs are shown in [Supplementary Tables A3–A5](#).

Significance was set at $p < 0.05$. All tests were two-tailed.

Interview—Health assessment time delay

A delay between the structured interview and health assessment is present. The median time delay between structured interview and health assessment was 56 days (IQR: 23 days – 109 days) (data not shown). Supplementary analyses were produced to assess whether delays between measurements contributed to discrepancies.

Results

Sample characteristics

Characteristics of the sample are presented in [Table 1](#). 53.6% of the sample were female. The sample had a mean age of 67.5 years (SD 9.1, range 50–94 years). 55.8% lived in an urban area. The majority of the sample were married (72.4%), reported that their health was excellent/very good/good (84.1%) and were retired or not working (73.4%). Over half the sample reported one or more chronic conditions (56.1%) and 48.1% reported one or more cardiovascular conditions. Just 8.4% reported use of sleep medication, while 9.3% reported

Table 1
Sample characteristics

	n = 1520
Sex, n (%)	
Male	704 (46.3)
Female	816 (53.7)
Age, Mean (SD) [Range]	67.5 (9.1) [50–94]
Education, n (%)	
Primary	387 (25.5)
Secondary	603 (39.7)
Tertiary	530 (34.9)
Marital Status, n (%)	
Married	1100 (72.4)
Never Married	117 (7.7)
Separated/Divorced	104 (6.8)
Widowed	199 (13.1)
Location, n (%)	
Urban	848 (55.8)
Rural	672 (44.2)
Employment Status ^a , n (%)	
Employed	403 (26.6)
Retired/Not Employed	1113 (73.4)
Self Rated Health, n (%)	
Excellent/Very Good/Good	1279 (84.1)
Fair/Poor	241 (15.9)
Medication use: Sleep, n (%)	128 (8.4)
Medication use: Anti-Hypertensives, n (%)	704 (46.3)
Medication use: Anti-Depressants, n (%)	142 (9.3)
Depressive Symptoms ^a , n (%)	137 (9.0)
Reported Pain ^a , n (%)	541 (35.7)
≥1 Chronic Condition, n (%)	852 (56.1)
≥1 Cardiovascular Condition, n (%)	731 (48.1)
Physical Activity Groups ^a , n (%)	
Low	521 (36.1)
Moderate	525 (36.4)
High	398 (27.6)
BMI Category ^a , n (%)	
Normal	351 (23.3)
Overweight	675 (44.9)
Obese	478 (31.8)
Current Smoker, n (%)	164 (10.8)
MMSE Score, mean (SD) [Range]	28.7 (1.7) [16–30]
Season of recording, n (%)	
Winter	294 (19.3)
Spring	252 (16.6)
Summer	461 (30.3)
Autumn	513 (33.8)

SD = Standard Deviation.

^a Missing data: employment status, n = 4 (0.3%); depressive symptoms, n = 5 (0.3%); reported pain, n = 4 (0.3%); physical activity groups, n = 76 (5.0%); BMI category, n = 16 (1.1%).

use of anti-depressants and 46.3% reported use of anti-hypertensive medication. The cohort were predominantly overweight (44.9%) or obese (31.8%). 9.0% reported depressive symptoms, while 35.7% reported the presence of pain. Only 10.8% of the sample were current smokers. The mean MMSE score for the sample was 28.7 (SD=1.7, range: 16–30).

Total sleep time

Table 2 summarises TSTsub and TSTobj by sample characteristics. Participants had a TSTsub of 7.0 hours (SD: 1.4, Range: 2.0–13.0 hours) of sleep on average compared to an TSTobj of 7.7 hours (SD: 1.2 hours, Range: 3.0–13.0 hours). The magnitude of TSTsub-TSTobj difference differed between demographics. Female participants had a sleep difference score of –0.8 h, while male participants had –0.6 hours. Those living in an urban area had a difference of –0.8 hours compared to –0.5 hours for those living in a rural area. The difference was –0.7 hours in Retired/Not Employed participants compared to –0.5 in employed participants. One of the largest

discrepancies of TSTsub and TSTobj was found in participants who were separated/divorced (–1.1 hours).

Differences in health status also had an impact on TSTsub-TSTobj difference. Those who reported fair/poor self-rated health had a significantly larger difference than those who reported excellent/very good/good self-rated health (–1.1 vs. –0.6 hours). Similarly, participants who used sleep medication had a score of –0.9 hours compared to –0.6 for those who were not taking sleep medication. The largest discrepancy was found in those who reported depressive symptoms who measured a difference of –1.7 hours, compared to –0.6 in those who did not report depressive symptoms.

Fig. 2 presents a Bland-Altman plot depicting TSTsub and TSTobj. The average agreement line between the two measures is shown to be below 0 (Observed Average Agreement = –0.68) suggesting systematic bias toward under-reporting of sleep compared to actigraphy-based sleep. The limits of agreement ranged from –3.90 to 2.55 hours.

Supplementary Table A.1 presents mean TSTsub-TSTobj differences by Interview – Health Assessment time delay quantiles. Those with the shortest delay between structured interview and health assessment (0–20 days) had a mean TSTsub-TSTobj difference of –0.73 hours (SD=1.74) compared to –0.70 (SD=1.69) in those with longer delays ranging between 77 and 127 days, and –0.52 (SD=1.60) in delays of between 128 and 421 days. No clear relationship between Interview-Health Assessment Time Delay was seen in a scatterplot depicting the two measures (Supplementary Fig. A.1). This is further assessed in Supplementary Table A.2 which presents a simple linear regression to model the association between TSTsub-TSTobj difference and Interview-Health Assessment time delay (log). The association was found to be non-significant [$B = 0.04$, 95% CI: –0.04, 0.12].

Scatterplots depicting average TSTsub-TSTobj difference and sleep problem components are shown in Fig. 3. Significant negative Spearman rank correlations were found between self-reported sleep and all sleep problem components suggesting that sleep difference scores became more negative, representing greater under-reporting of self-reported sleep, as sleep problems increased. Correlations were strongest with trouble waking too early ($R_s = -0.454$, $p < 0.001$) and trouble falling asleep ($R_s = -0.345$, $p < 0.001$). Only daytime sleepiness was significantly correlated with actigraphy-based sleep, however this correlation was weak ($R_s = -0.082$, $p < 0.01$).

The ICC coefficient between TSTsub and TSTobj was 0.18 (Table 3). The highest ICC coefficients were measured in those who reported trouble waking too early “rarely or never” (ICC=0.29) and trouble falling asleep “rarely or never” (ICC=0.26). The lowest coefficients were obtained in measurements of participants who reported trouble falling asleep or trouble waking up too early “most of the time”, with ICC coefficients of 0.03 and 0.02 respectively.

Self-reported and actigraphy-based sleep differences

Fig. 4 depicts mean TSTsub and TSTobj by sleep problem measures. Sleep difference scores remained stable for those reporting increased chances of daytime sleepiness, with those reporting that they would never doze, and those reporting a high chance of dozing both measuring a TSTsub-TSTobj difference of –0.8 hours (Fig. 4A). TSTsub-TSTobj differences were lowest in those who reported “rarely or never” having trouble falling asleep (–0.3 hours) (Fig. 4B), or trouble waking too early (–0.2 hours) (Fig. 4C). These differences increased as the frequency of either problem increased. Participants who stated that they have trouble falling asleep “most of the time” had a difference of –2.3 hours. Those who stated they have trouble waking too early “most of the time” had a difference of –2.2 hours.

Regression models predicting sleep difference scores for individual sleep problem components are shown in Fig. 5. Basic models and

Table 2

Sample characteristics by self-reported total sleep time, actigraphy-based total sleep time and self-reported total sleep time – actigraphy-based total sleep time difference

	Self-reported total sleep time (h)	Actigraphy-based total sleep time (h)	TSTsub-TSTobj difference (h)
Overall	7.0	7.7	–0.7
Sex			
Male	7.1	7.7	–0.6
Female	7.0	7.8	–0.8
Education			
Primary	7.0	7.8	–0.8
Secondary	7.0	7.7	–0.7
Tertiary	7.1	7.7	–0.5
Marital Status			
Married	7.0	7.7	–0.7
Never Married	7.0	7.5	–0.5
Separated/Divorced	6.8	7.8	–1.1
Widowed	7.2	7.8	–0.6
Location			
Urban	6.9	7.7	–0.8
Rural	7.2	7.7	–0.5
Employment Status ^a			
Employed	7.0	7.5	–0.5
Retired/Not Employed	7.1	7.8	–0.7
Self Rated Health			
Excellent/Very Good/Good	7.1	7.7	–0.6
Fair/Poor	6.6	7.7	–1.1
Sleep Medication: No	7.0	7.7	–0.6
Sleep Medication: Yes	6.9	7.9	–0.9
Anti-Hypertensive Medication: No	7.0	7.7	–0.6
Anti-Hypertensive Medication: Yes	7.1	7.8	–0.7
Anti-Depressant Use: No	7.0	7.7	–0.7
Anti-Depressant Use: Yes	7.2	7.9	–0.8
Depressive Symptoms: No	7.1	7.7	–0.6
Depressive Symptoms: Yes	6.2	7.8	–1.7
Reported Pain: No	7.2	7.7	–0.5
Reported Pain: Yes	6.8	7.7	–1.0
≥1 Chronic Condition: No	7.0	7.7	–0.6
≥1 Chronic Condition: Yes	7.0	7.8	–0.7
≥1 Cardiovascular Condition: No	7.1	7.7	–0.6
≥1 Cardiovascular Condition: Yes	7.0	7.8	–0.8
Physical Activity Groups			
Low	7.0	7.7	–0.7
Moderate	7.1	7.8	–0.8
High	7.1	7.7	–0.5
BMI Category			
Normal	6.9	7.6	–0.7
Overweight	7.1	7.7	–0.7
Obese	7.1	7.7	–0.6
Current Smoker: No	7.0	7.7	–0.7
Current Smoker: Yes	7.0	7.8	–0.8
Season of recording			
Winter	7.1	7.9	–0.8
Spring	7.1	7.6	–0.5
Summer	7.1	7.6	–0.5
Autumn	7.0	7.8	–0.8

TSTobj = Actigraphy-based total sleep time; TSTsub = Self-Reported total sleep time.

fully adjusted models are displayed. Positive coefficients represent over-reporting of sleep, negative coefficients represent under-reporting of sleep.

In the fully adjusted models, only participants who reported a slight chance of daytime sleepiness compared to no chance had a significant positive association with sleep difference scores ($B = 0.31$, 95% CI: 0.08,0.51, $p < 0.01$) (Fig. 5A).

When compared to participants who stated they “rarely or never” had trouble falling asleep, or waking up too early, those who reported they had trouble either “sometimes” or “most of the time” had significant negative associations with sleep difference scores. The effects were attenuated in fully adjusted models, but remained significant. Strongest associations were found in those who had trouble falling

asleep “most of the time” ($B = -1.80$ 95% CI: $-2.10, -1.51$, $p < 0.001$) (Fig. 5B) compared to trouble falling asleep “rarely or never”, or trouble waking up too early “most of the time” ($B = -1.86$ 95% CI: $-2.10, -1.63$, $p < 0.001$) (Fig. 5C) when compared to trouble waking up too early “rarely or never”.

The r^2 value for the basic Daytime Sleepiness model was just 0.01, suggesting that although significant associations were found, Daytime Sleepiness predicts little of the variance (1%) in sleep reporting scores (Supplementary Table A3). In contrast, the r^2 value for the basic model using trouble falling asleep to predict variance in sleep difference scores was 0.18 (Supplementary Table A5). This increased only slightly to 0.21 where all covariates were included in the fully adjusted model.

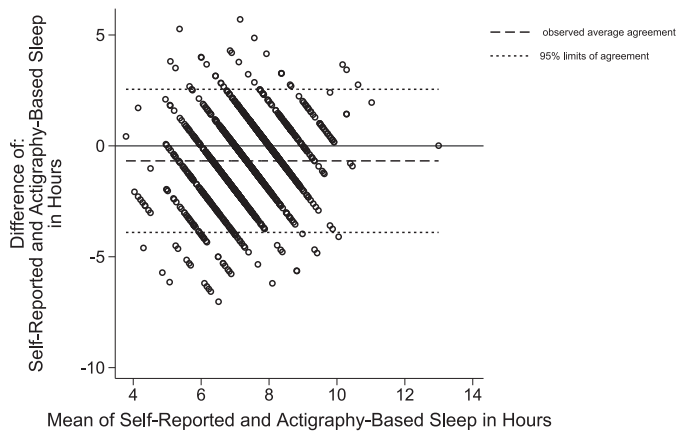


Fig. 2. Bland-Altman plot comparing self-reported and actigraphy-based total sleep time. Actigraphy-based total sleep time is presented using measurements prior to rounding.

Discussion

This study sought to assess measurement agreement between self-reported total sleep time and actigraphy-based total sleep time using data from TILDA. Overall agreement was low between measurements with systematic under-reporting of sleep relative to actigraphy-based sleep apparent. Average results from both measurements were within one hour of each other, but participants under-reported their sleep by 0.7 hours compared to actigraphy-based sleep. Differences were more pronounced in participants reporting poorer health, such as fair/poor self-rated health and depressive symptoms, with discrepancies of greater than one hour between measurements.

Another aim of this study was to determine the impact of subjective sleep problems on measurement agreement using three short sleep problem questions. Sleep problems were prevalent. Over one-third of participants experienced insomnia symptoms which is in accordance with previous literature.²¹ Agreement between measurements decreased in those experiencing insomnia symptoms as the magnitude of sleep under-reporting increased. In particular, those who experienced trouble falling asleep, or waking too early “most of the time” were found to under-report their sleep by over two hours

on average. This is consistent with other findings which showed that poor self-reported sleep quality was associated with shorter reported TST when compared to measured sleep.^{11,12,14} Experience of daytime sleepiness did not show the same effects on sleep reporting in this sample. Measurement differences were consistent in all categories which may suggest daytime sleepiness is not a driving force to the same extent of insomnia symptoms in perceptions of sleep. Use of the TILDA data in this analysis enabled extensive adjustment for relevant confounders in assessing these questions as driving factors of these discrepancies.

Depressive symptoms were shown to be the strongest driving participant characteristic in sleep reporting differences. To further understand this relationship, supplementary analyses assessing prevalence of depressive symptoms and anti-depressant use by sleep problems were conducted. Depressive symptoms were more prominent in participants reporting greater magnitude of sleep problems (Supplementary Table A6). Similarly, use of antidepressant medication was more commonly reported in those reporting more sleep problems (Supplementary Table A7). This is consistent with previous findings of the association between insomnia and depressive symptoms.^{2,6}

Accelerometer devices may classify a period of sedentary, motionless behaviour as sleep. They are still however considered an effective method for sleep research.¹⁶ We have shown that in a sample of older adults, actigraphy-based measurements were on average longer than self-reported sleep.

Underreporting was most pronounced in participants who self-reported insomnia symptoms. It is suggested that discrepancies of this kind in insomniacs may be a reflection of poor sleep quality that is not captured by accelerometer devices rather than a misperception.¹¹ It is unclear whether differences here were under-reporting sleep relative to how much sleep was actually obtained, or whether recordings were longer resulting from sedentary periods resulting from trouble falling asleep or waking too early. The day-to-day variability of sleep duration experienced in those with insomnia complaints creates challenges in using either objective measurement or single-item survey questions. With consideration to the high prevalence of insomnia in the older population,²¹ researchers should be cognisant of the impact this may have on either self-reported or actigraphy-based measurement.

These discrepancies have implications on studies of sleep duration and it is recommended where possible that multiple measures of sleep are included.^{9,11,22} Landry et al. note the importance of the

Table 3

Self-reported total sleep time – actigraphy-based total sleep time difference, intra-class correlations and bland-altman 95% limits of agreement results overall self-reported and actigraphy-based total sleep time and by daytime sleepiness, trouble falling asleep, trouble waking too early

	n (%)	TSTsub-TSTobj difference	ICC	Bland altman 95% limits of agreement	
Overall		−0.7	0.18	−2.60	3.95
Daytime Sleepiness					
Would never doze	591 (38.9)	−0.8	0.17	−2.54	4.15
Slight chance of dozing	431 (28.4)	−0.5	0.17	−2.53	3.46
Moderate chance of dozing	245 (16.1)	−0.6	0.17	−2.63	3.78
High chance of dozing	253 (16.6)	−0.8	0.19	−2.70	4.40
Trouble falling asleep					
Rarely or never	1000 (65.5)	−0.3	0.26	−2.63	3.32
Sometimes	379 (24.9)	−0.9	0.15	−2.19	4.01
Most of the time	149 (9.6)	−2.3	0.03	−1.28	5.91
Trouble waking too early					
Rarely or never	800 (52.2)	−0.2	0.29	−2.68	3.01
Sometimes	489 (32.1)	−0.8	0.20	−2.12	3.65
Most of the time	243 (15.7)	−2.2	0.02	−1.26	5.69

TSTobj = Actigraphy-based total sleep time; TSTsub = Self-reported total sleep time; ICC = Intra-Class Correlation.

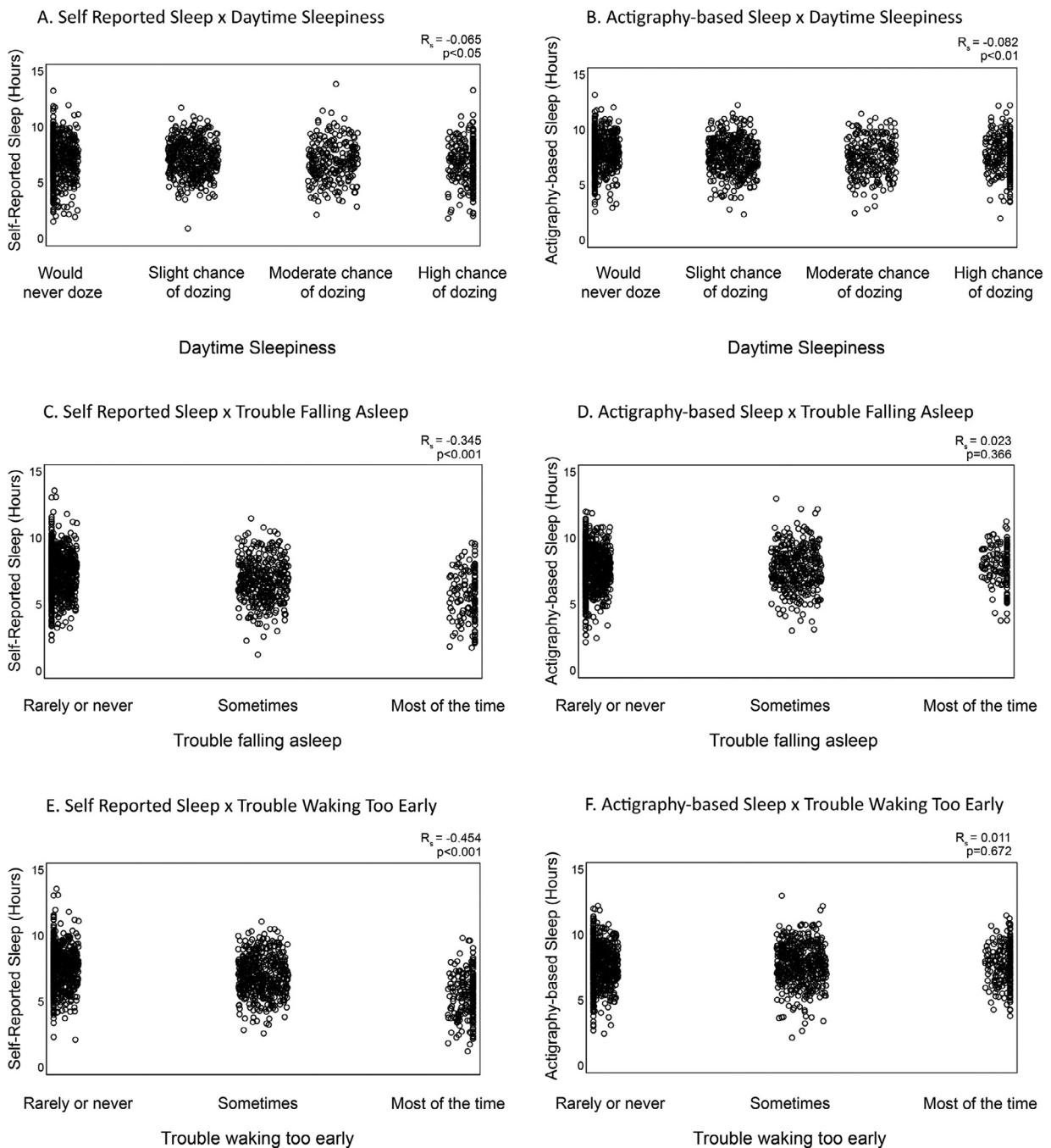


Fig. 3. Scatterplots of average difference between self-reported and actigraphy-based total sleep time in hours by Daytime sleepiness, Trouble falling asleep and Trouble waking too early.

methods used in sleep quality assessment given the complex nature of sleep.⁹ The PSQI has been commonly used as the method of measuring sleep quality in research of this kind,^{9-12,19,28} but the length of this tool may be prohibitive in certain settings. Our findings demonstrate that addition of short, quick to administer questions which assess insomnia symptoms may be effective complimentary questions to account for potential bias when evaluating sleep duration. These results add to current literature with this evidence, and confirm findings of a similar nature using a large, population-derived cohort of older adults.

These findings also have implications for clinical practice. Those with adverse health characteristics were shown to be most at risk of discrepancies. Measurement on these cohorts should emphasise use

of objective measurement using wearable devices such as FitBit or similar.

Strength and limitations

This study has a number of strengths. This is one of the largest studies of its kind to date and uses a population derived community-dwelling cohort of older adults. The design of TILDA allowed for an assessment of agreement of sleep measurements in older adults while accounting for a comprehensive array of confounding factors.

Limitations are also present. The accelerometer measurements are from a sub-sample of the TILDA study and not population representative, however it has been shown that this sample is characteristically

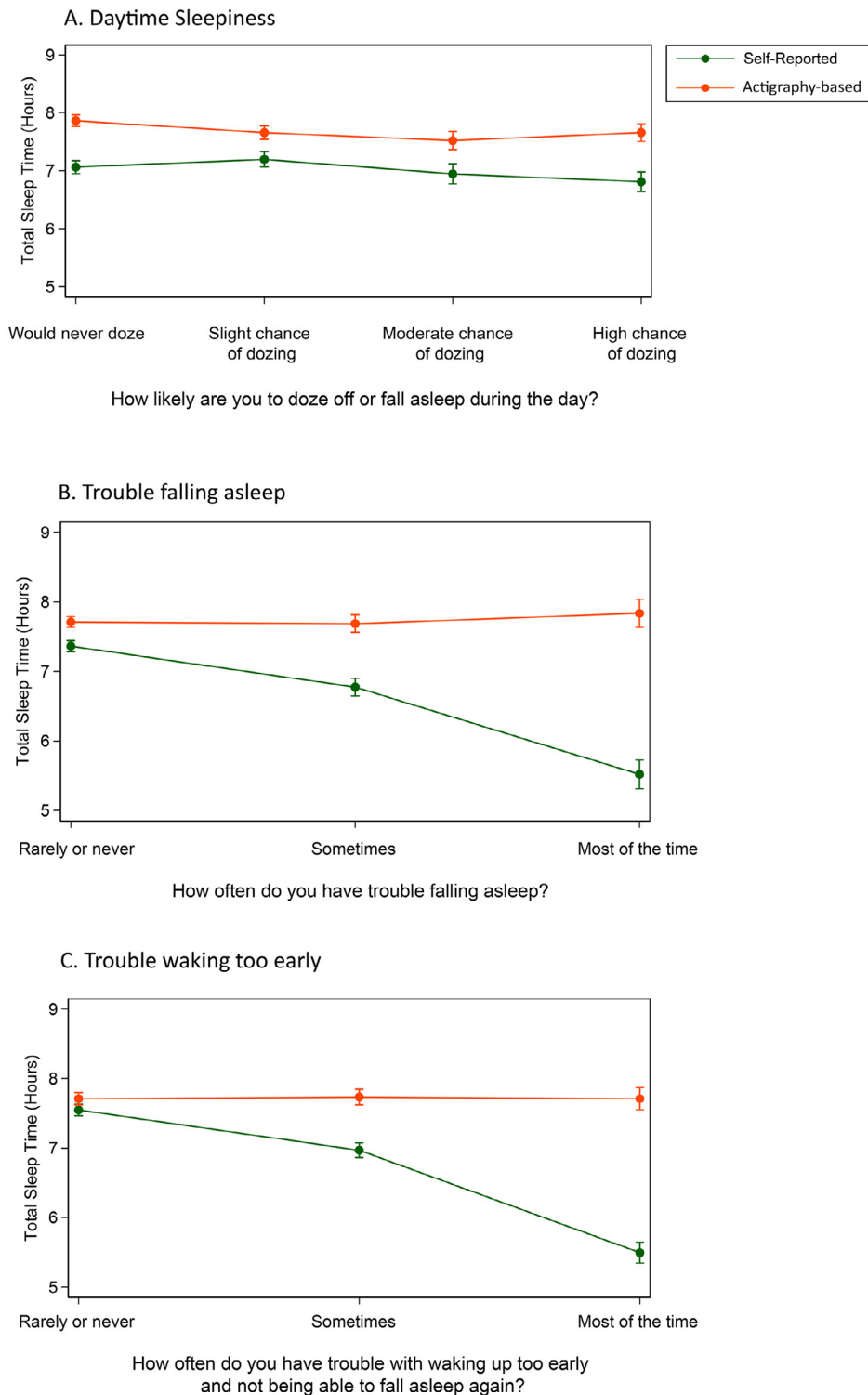


Fig. 4. Means and 95% confidence intervals of Self-Reported Total Sleep Time and Actigraphy-based Total Sleep Time by (A) Daytime Sleepiness, (B) Trouble Falling Asleep, and (C) Trouble waking too early.

similar to the full cohort.³⁴ Participants are asked when reporting their sleep to round to the nearest hour, reducing the precision of this measurement. Self-Reported Total Sleep Time was asked during the self-interview, while accelerometer measurements were taken directly after the health assessment. As a consequence of the study design of TILDA, a delay between the structured interview and health

assessment is present. This created some uncertainty in whether reported sleep reflected the same sleep patterns that were being experienced during the accelerometer recording period. We produced supplementary analyses to assess whether this contributed to measurement discrepancies. We did not find clear evidence for an effect resulting from the lag. Nevertheless, lag between assessments

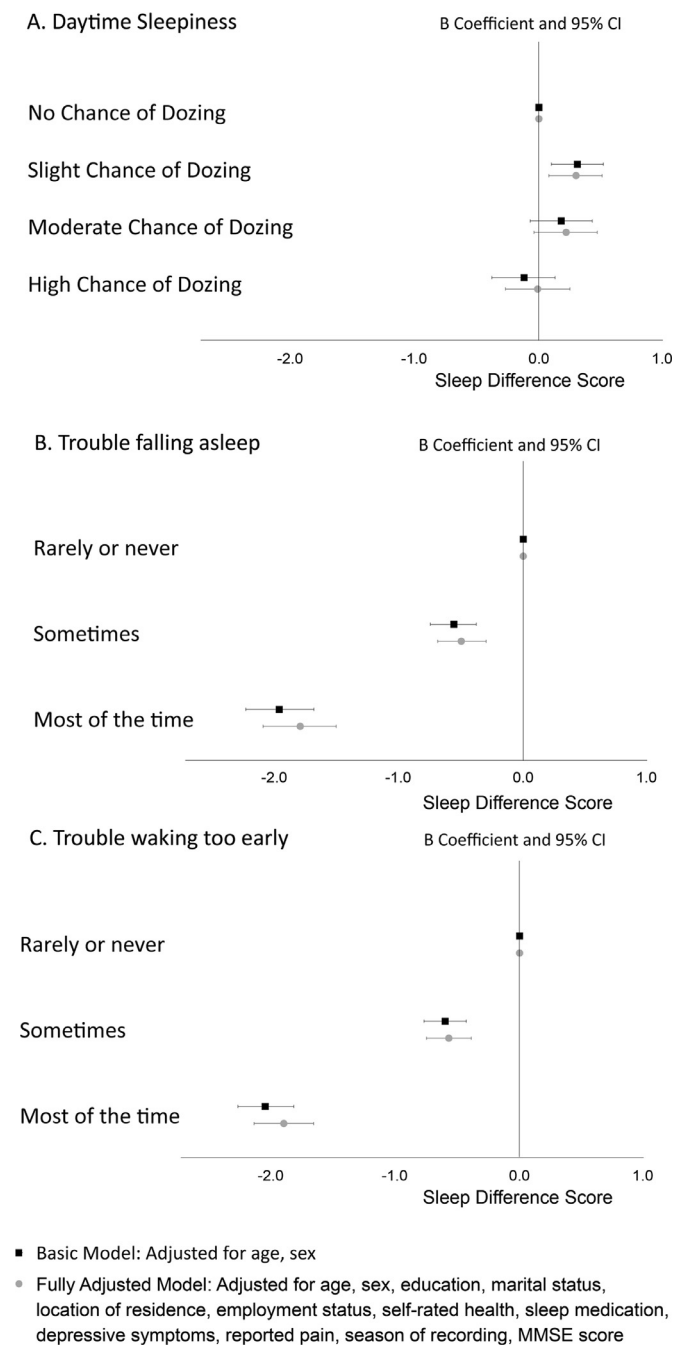


Fig. 5. Linear regression models for predictors of TSTsub-TSTobj difference by (A) Daytime Sleepiness, (B) Trouble Falling Asleep, and (C) Trouble waking too early. CI, confidence interval. Positive coefficients represent over-reporting of sleep, negative coefficients represent under-reporting of sleep. TSTobj = Actigraphy-based total sleep time; TSTsub = Self-reported total sleep time.

is a potential limitation in studies of variable characteristics like sleep duration. Studies of a similar nature should be conscious of potential contributions to discrepancies where delays between measurements are present.

Conclusions

In summary, agreement between self-reported and actigraphy-based sleep duration was low in this sample of community dwelling older adults. Health practitioners should be aware of the prevalence of under-reporting of sleep duration, particularly as those with

adverse health characteristics were shown to be most at risk of discrepancies. Insomnia symptoms were associated with measurement discrepancies. It was shown that this effect can be gauged using quick to administer questions. Studies which seek to use either form of measurement should consider inclusion of similar questions as a means of establishing whether self-reporting, or objective measurement may have been influenced. Future work will investigate longitudinal differences in measurements and incident insomnia symptoms to better understand the directionality of the relationship.

Declaration of conflict of interest

The authors declare no conflict of interest.

Acknowledgments

The authors would like to thank The Irish Longitudinal Study on Ageing (TILDA) study participants, research team, study nurses, and administrative staff. TILDA is funded by the Department of Health, Irish Life, and The Atlantic Philanthropies. Researchers interested in using TILDA data may access the data for free from the following sites: Irish Social Science Data Archive at University College Dublin (<http://www.ucd.ie/issda/data/tilda/>). Interuniversity Consortium for Political and Social Research at the University of Michigan (<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/34315>).

Supplementary materials

Supplementary material associated with this article can be found in the online version at <https://doi.org/10.1016/j.sleh.2020.06.003>.

References

1. Yaffe K, Falvey CM, Hoang T. Connections between sleep and cognition in older adults. *Lancet Neurol.* 2014;13:1017–1028.
2. Alvaro PK, Roberts RM, Harris JK. A systematic review assessing bidirectionality between sleep disturbances, anxiety, and depression. *Sleep.* 2013;36:1059–1068.
3. Cribbet MR, Carlisle M, Cawthon RM, et al. Cellular aging and restorative processes: subjective sleep quality and duration moderate the association between age and telomere length in a sample of middle-aged and older adults. *Sleep.* 2014;37:65–70.
4. Spira AP, Chen-Edinboro LP, Wu MN, Yaffe K. Impact of sleep on the risk of cognitive decline and dementia. *Curr Opin Psychiatry.* 2014;27:478–483.
5. Spiegelhalter K, Scholtes C, Riemann D. The association between insomnia and cardiovascular diseases. *Nat Sci Sleep.* 2010;2:71–78.
6. Roberts RE, Shema SJ, Kaplan GA, Strawbridge WJ. Sleep complaints and depression in an aging cohort: a prospective perspective. *Am J Psychiatry.* 2000;157:81–88.
7. Gallicchio L, Kalesan B. Sleep duration and mortality: a systematic review and meta-analysis. *J Sleep Res.* 2009;18:148–158.
8. Lee CH, Sibley CG. Sleep duration and psychological well-being among New Zealanders. *Sleep Health.* 2019;5:606–614.
9. Landry G, Best J, Liu-Ambrose T. Measuring sleep quality in older adults: a comparison using subjective and objective methods. *Front Aging Neurosci.* 2015;7:1–10.
10. Schokman A, Bin YS, Simonelli G, et al. Agreement between subjective and objective measures of sleep duration in a low-middle income country setting. *Sleep Health.* 2018;4:543–550.
11. Van Den Berg JF, Van Rooij FJ, Vos H, et al. Disagreement between subjective and actigraphic measures of sleep duration in a population-based study of elderly persons. *J Sleep Res.* 2008;17:295–302.
12. Hughes JM, Song Y, Fung CH, et al. Measuring sleep in vulnerable older adults: a comparison of subjective and objective sleep measures. *Clin Gerontol.* 2018;41:145–157.
13. Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? *Epidemiology.* 2008;19:838–845.
14. McCrae CS, Rowe MA, Tierney CG, Dautovich ND, Definis AL, McNamara JP. Sleep complaints, subjective and objective sleep patterns, health, psychological adjustment, and daytime functioning in community-dwelling older adults. *J Gerontol B Psychol Sci Soc Sci.* 2005;60:P182–P189.
15. Kurina LM, Thisted RA, Chen J-H, McClintock MK, Waite LJ, Lauderdale DS. Actigraphic sleep characteristics among older Americans. *Sleep Health.* 2015;1:285–292.
16. Sadeh A, Acebo C. The role of actigraphy in sleep medicine. *Sleep Med Rev.* 2002;6:113–124.
17. Means MK, Edinger JD, Glenn DM, Fins AI. Accuracy of sleep perceptions among insomnia sufferers and normal sleepers. *Sleep Med.* 2003;4:285–296.

18. Girschik J, Fritschi L, Heyworth J, Waters F. Validation of self-reported sleep against actigraphy. *J Epidemiol*. 2012;22:462–468.
19. Jackowska M, Ronaldson A, Brown J, Steptoe A. Biological and psychological correlates of self-reported and objective sleep measures. *J Psychosom Res*. 2016;84:52–55.
20. National Sleep Foundation. *How Much Sleep Do Adults Need?* Washington DC: National Sleep Foundation; 2010. Available at: <http://www.sleepfoundation.org/article/white-papers/how-much-sleep-do-adults-need.15/11/2019>. Accessed 15 November 2019.
21. Patel D, Steinberg J, Patel P. Insomnia in the elderly: a review. *J Clin Sleep Med*. 2018;14:1017–1024.
22. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. *Sleep Med Rev*. 2011;15:259–267.
23. Buysse 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.
24. Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep*. 1991;14:540–545.
25. Roenneberg T, Wirz-Justice A, Mrosovsky M. Life between clocks: daily temporal patterns of human chronotypes. *J Biol Rhythms*. 2003;18:80–90.
26. van Hees VT, Sabia S, Anderson KN, et al. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. *PLoS ONE*. 2015;10: e0142533.
27. Kenny RA, Whelan BJ, Cronin H, et al. The design of the Irish longitudinal study on ageing. Dublin: The Irish Longitudinal Study on Ageing; 2010. <https://doi.org/10.38018/TildaRe.2010-00>.
28. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort profile: the English longitudinal study of ageing. *Int J Epidemiol*. 2012;42:1640–1648.
29. Lauderdale DS, Knutson KL, Yan LL, et al. Objectively measured sleep characteristics among early-middle-aged adults: the CARDIA study. *Am. J. Epidemiol*. 2006;164:5–16.
30. Kearney PM, Cronin H, O'Regan C, et al. Cohort profile: the Irish Longitudinal Study on Ageing. *Int J Epidemiol*. 2011;40:877–884.
31. Whelan B. A national random sampling design for Ireland. *Econ Soc Rev*. 1979;10:169–174.
32. McGarrigle C, Donoghue O, Scarlett S, Kenny RA. Health and wellbeing: active ageing for older adults in Ireland. *Evidence from The Irish Longitudinal Study on Ageing. Dublin: The Irish Longitudinal Study on Ageing*. 2017. <https://doi.org/10.38018/TildaRe.2017-01>.
33. Donoghue OA, McGarrigle CA, Foley M, Fagan A, Meaney J, Kenny RA. Cohort profile update: the Irish Longitudinal Study on Ageing (TILDA). *Int J Epidemiol*. 2018;47:1398–1398.
34. Scarlett S., Nolan H., Kenny RA., O'Connell M.D. Objective Sleep Duration in Older Adults: results from The Irish longitudinal study on ageing. *J Am Geriatr Soc*. Doi:10.1111/jgs.16177.
35. Nolan H, Dunne CL, Greene L, et al. Signal processing framework for objective sleep and physical activity measurement using accelerometry in population studies of older adults. *Age Ageing*. 2013;42. iii15.
36. Jenkins CD, Stanton B-A, Niemcryk SJ, Rose RM. A scale for the estimation of sleep problems in clinical research. *J Clin Epidemiol*. 1988;41:313–321.
37. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189–198.
38. Briggs R, Carey D, O'Halloran AM, Kenny RA, Kennelly SP. Validation of the 8-item centre for epidemiological studies depression scale in a cohort of community-dwelling older people: data from The Irish Longitudinal Study on Ageing (TILDA). *Eur Geriatr Med*. 2018;9:121–126.
39. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.
40. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.