Title: Are sleep quality judgements comparable across individuals, places and spaces? An interdisciplinary analysis of data from 207,608 individuals across 68 countries.

Authors: Robert Meadows PhD[1], Ian Brunton-Smith PhD[1] and Jason Ellis PhD[2]

1 Department of Sociology, University of Surrey, Surrey, UK
2 Department of Psychology, Northumbria University, Newcastle, UK

Corresponding author: Professor Robert Meadows, Department of Sociology, University of Surrey, Guildford, Surrey, GU27XH Email: r.meadows@surrey.ac.uk

Word count: 3922 (not including abstract, acknowledgement, references and figures)

Abstract

Objectives: Significant research has shown that health is a heterogeneous concept, and one person's poor health may not be comparable with another's. Yet, little consideration has been given to whether sleep quality judgements are also heterogenous or whether they cohere between individuals. Another possibility is that there are group differences in the ways in which sleep quality is perceived. If this is the case, it is possible known inequalities in sleep are – in part – an artefact of social position influencing how we conceive of sleep problems. The current study explores this possibility.

Design: Cross-sectional, using World Health Organisation data from 207,608 individuals; aged between 15 to 101 years of age from 68 countries. Alongside a battery of sleep and demographic variables, data contained sleep and energy vignettes. Random effect anchoring vignette models were applied to investigate interpersonal incompatibility and whether sleep quality perceptions operate differently depending on social location, context and function.

Results: Whilst sleep quality judgements are largely comparable across individuals, findings also highlight how the relationship between education and self-reported sleep changes following adjustment for reporting heterogeneity. Estimates of threshold parameters suggest that those with more years of education have a slightly increased threshold for reporting mild sleep problems (B 0.005; s.e. 0.001) but a lower threshold for reporting sleep problems as extreme (B -0.007; s.e. 0.001).

Conclusions: Sleep quality judgements occupy a complex position between heterogeneity and coherence. This has implications for both epidemiological methodologies and contemporary debates about social justice, public health and sleep.

Keywords: sleep quality, boundary object, inequalities, education, global

Introduction

'Sleep quality' is widely recognized as an important indicator of health and wellbeing; as well as an important patient reported outcome used to determine both treatment need and success [1]. It is associated with positive outcomes such as better health and wellbeing and [2] in its negative form, is one of the defining features of chronic insomnia [2] and a key predictor of CVD [3, 4].

Unsurprisingly, considerable attention has been paid to understanding sleep quality. An early concern was with *definition* and *measurement*. In the late 1980s, Buysse et al [5] noted that sleep quality is a "complex phenomenon that is difficult to define and measure objectively." A few years later, Akerstedt et al [6] noted that "there seems to be very little systematic knowledge as to what actually constitutes subjectively good sleep and how this should be measured". Some 20 years after Buysse et al's comment Ohayon et al [7] suggested that:

> "[A]lthough, the term "sleep quality" is widely used by researchers, clinicians, and the public, this expression lacks definitional consensus. To date, no consistent guidance is available from the scientific community regarding what constitutes normal or optimal, healthy sleep and good sleep quality."

This concern with definition and measurement has led to considerable empirical attention being paid to the *determinants* of sleep quality. Research has illustrated how sleep quality judgments are influenced by sleep continuity, ease of awakening, reported dreaming [6] and both memory of night-time sleep and daytime functioning [1]. Other important elements include how individuals feel about the bed and the bedroom [8]. Whilst there are more similarities than differences between insomnia and normal sleeper groups in the meaning of sleep quality, those experiencing insomnia consider more things important for judging sleep quality relative to 'normal sleepers' [2].

There are, however, complexities to this idea of *determinants.* As Ogeil and colleagues [9] highlight, often the same factors are implicated in both good and bad night's sleep and people often outline similar sleep environments when describing both types of night. What is important, these authors argue, is how factors coalesce and are ‘configured’ and ‘assembled’ in any given situation. For example, alcohol and drugs can impact sleep quality judgements in both positive and negative ways depending on how they combine with other factors, such as activities preceding sleep. This points towards a further question. In all of the examples above, the focus is on how *levels* of X variable influence *levels* of reported sleep quality (whether alone or in combination). This assumes that there is a single, coherent, concept of sleep quality, which is in operation. Yet, there are reasons to question this. Taking the example of ‘health’ more generally, there is extensive literature suggesting health judgments may be ‘syncretic in origin’, derived from disparate and distinct sources and may vary between social, geographical and cultural contexts [10-12]. Conceptualizations of sleep quality may operate similarly and rather than asking ‘what are the determinants of sleep quality?’ we need to ask questions such as ‘what is sleep quality’ and ‘does it operate in different ways for different people’?

Generating hypotheses: From boundary objects to differential item functioning

Different disciplines have highlighted the importance of asking these questions; yet this has often been in subtle, unacknowledged, or incomplete ways.   The clinical literature has largely engaged with these questions under the premise of an objective/subjective divide; for example evaluating the gap between subjective and objective measures of sleep [13], exploring which objective markers might be the best predictor of subjective sleep quality [14], or attempting to reach consensus through systematic reviews and modified DELPHI surveys [7]. Sociological studies have also explored the ways in which people engage with sleep (quality); arguing, for example, that there are four modes of embodied sleep – namely the normative, pragmatic, visceral and experiential – and, at any given time, individuals will prioritize one mode over the other.As Meadows [15] writes, “an individual who avoids alcohol at night because it is said to damage sleep engages with the normative; whereas an individual who avoids alcohol at night because it damages sleep and *this is* *seen as problematic because it disrupts everyday functions, such as being a parent*, prioritises the pragmatic" [15].

Whilst the idea of complete heterogeneity in sleep quality judgements appears to be negated within this literature, the opaque evidence base does still generate conflicting hypotheses. It has, for example, been suggested that sleeplessness – and insomnia in particular – are *boundary objects which cohere across people and spaces [16]*. The concept of 'boundary object' was originally developed to understand a 'central tension' in science between heterogeneity and cooperation [17]. It describes artifacts and terminologies that are used to translate knowledge, build consensus and encourage cooperation [18]; such as standardized forms and ideal types. These boundary objects can have different meanings in different social worlds but are sufficiently structured to be recognised across them [19]. Insomnia, for example, oscillates between 'normality' and 'pathology' and between symptom and disorder. Whilst diagnostic systems and guidelines exist [20], insomnia still displays an element of 'interpretive flexibility' within and across different sites and settings [16]. It is possible, then, that sleep quality judgements are also boundary objects – "both 'plastic enough' to 'travel' across different boundaries, sites and settings and 'robust enough' to maintain some sort of coherence in so doing, albeit 'without consensus' [16].

Literature also suggests an alternative hypothesis and that there are group differences in the ways in which sleep quality is perceived. Within sleep science, debates exist surrounding the nature of discrepancies in objective and subjective sleep quality both within and between genders [21-25]. Current research on inequalities in sleep also highlights how race, ethnicity, culture, employment, neighbourhood, socioeconomic status, marriage and the family environment can all impact on an individual's sleep [26]. Sociologists have suggested that these differences may be linked to reporting thresholds as there are "contingent influences on the meanings attached to sleep" and the ways in which sleep is "commonsensically conceived will relate to an individual's social location and economic function" [27]. As Taylor [27] goes on to argue, those who see sleep as a leisure pursuit are liable to have attained a certain socio-economic status, whereas 'hard-working peasantry' are more likely to view sleep as respite from labour. Ideas of 'group difference' can also be seen in the survey methodological literature, albeit from a different conceptual frame. In an examination of 'reporting bias', Huang and Kampfton [18] suggest that individuals with depressive symptoms might "report poor sleep quality because they are indeed deprived of sleep as a result of depression, or simply because they are inclined to rate everything more negatively, making subjective sleep quality assessments between individuals with and without depressive symptoms hard to compare" [28]. In the latter framework, depression is said to impact on the ways in which we conceive of sleep quality. Depression may, for example, lead to individuals interpreting the question differently or interpreting the response scale differently, which is more formally known as differential item functioning (DIF) and individual reporting heterogeneity.

This brief review highlights how numerous disciplines have asked how sleep quality judgements operate without overtly exploring this. It also demonstrates how they approach the issue differently and in silos. The present study merges ideas from across the disparate literatures described above; using analytical techniques developed to identify and adjust for DIF to explore whether sleep quality judgments cohere or whether they vary depending on context, social location and function.

Method

This is an observational, cross-sectional, study. The study draws on World Health Organisation data from 207,608 individuals within 68 countries and applies random effect anchoring vignette models to explore whether there is interpersonal incompatibility across sleep questions. As Tareque et al [29] highlight, perceptions of health are based on latent continuous scales and individuals often have different 'cut points' based on personal background and experience (i.e. DIF). What this means in practice, for example, is that the same level of health may be considered 'poor' by one person and 'adequate' by another. Anchoring vignettes are often used to account for DIF. These are "hypothetical descriptions of various levels of health". The idea is that respondents are presented with different scenarios such as the following: “[Kevin] suffers from back pain that causes stiffness in his back, especially at work, but is relieved with low doses of medication. He does not have any pains other than this generalised discomfort” [30]. Participants are then asked to rate the severity of the problem, with each participant providing ratings for a number of scenarios capturing different levels of problem severity. These ratings are then used as an anchor to adjust participant self-reported assessments of their own health. Levels of the vignette are invariant over respondents [31], so any variability in self-reports would reveal ‘true’ differences in underlying health.

This methodology rests on a set of assumptions. As Tareque et al [29] write “First, irrespective of any characteristic, all respondents perceive the level of health represented in each vignette the same way, only assuming random measurement error (vignette equivalence). Second, respondents use the same response scale and set of cut-points to evaluate all vignettes and a self-assessment question for each domain (response consistency).” Whilst primarily used to explore whether we need to adjust for reporting heterogeneity, these models also allow for substantive investigation of interpersonal incompatibility and therefore enable us to explore the level to which individual, demographic and cultural differences exist in perceptions of sleep quality.

Data

Sleep and energy vignettes formed part of the World Health Survey Study that was implemented between 2002 and 2004 to “strengthen national capacity to monitor critical health outcomes and health systems through the fielding of valid, reliable and comparable household survey instrument”. It aimed to capture data from all regions of the world. Whilst these have been analysed previously, this has largely been focused on the Japanese data [29]. The survey was fielded to a total of 275,272 individuals across 69 countries

The World Health Survey was conducted under the supervision of the WHO. There was an independent ethics review of the protocol and informed consent was obtained from respondents.

This particular analysis was exempt from further full review as it was solely based on anonymous, non-identifiable, data [32]

Measures

Sleep quality is measured with both a single self-rated sleep item, and 6 anchoring vignettes. The self-rated sleep item asks: “Overall in the last 30 days, how much of a problem did you have with sleeping, such as falling asleep, waking up frequently during the night or waking up too early in the morning?”, with response options “None”, “Mild”, “Moderate”, “Severe” and “Extreme”

The anchoring vignettes asked respondents to rate the following series of scenarios (ordered from least to most severe) on the same 5-point scale, with the order of the scale randomised.

- [Mark] falls asleep every night within five minutes of going to bed. He sleeps soundly during the whole night and wakes up in the morning feeling well-rested and feels full of energy all day.
- [Paolo] has no trouble falling asleep at night and does not wake up during the night, but every morning he finds it difficult to wake up. He uses an alarm clock but falls back asleep after the alarm goes off. He is late to work on four out of five days and feels tired in the mornings.
- [Noemi] falls asleep easily at night, but two nights a week she wakes up in the middle of the night and cannot go back to sleep for the rest of the night. On these days she is exhausted at work and cannot concentrate on her job.
- [Damien] wakes up almost once every hour during the night. When he wakes up in the night, it takes around 15 minutes for her to go back to sleep. In the morning she does not feel well-rested and feels slow and tired all day.
- [Daniel] takes about two hours every night to fall asleep. He wakes up once or twice a night feeling panicked and takes more than one hour to fall asleep again. Three to four nights a

week he wakes up in the middle of the night and cannot go back to sleep for the rest of the night. He is fatigued all day, every day and misses work several times a week. He cannot take part in sports or social activities.

Initial analysis across countries confirms an increasing tendency for respondents to rate sleep problems more severely when moving from the first to the fifth vignette scenario.

Independent variables were selected to explore whether sleep quality is conceived of differently depending on social location and function. Variables included gender (male/female), age (continuous), employed/not employed, marital status (married/cohabitating, separated/divorced, widowed and single), years in formal education (continuous) and health status. Self-rated health asked "in general how would you rate your health today? (1 = very good, 2 = good, 3 = moderate, 4 = bad, 5 = very bad). Following Tareque et al [29], number of comorbidities was generated from data on the presence of seven diseases including angina, chest pain, asthma, wheezing, feeling sad, empty or depressed and reported loss of interest in most enjoyable things.

Missing data

Whilst missing data on self-rated health was minimal (n=3,824, 1.4%), there was a larger proportion of missing data for employment status (n = 38569) and years in education (n=29586), with the remainder of variables included showing less than 5% missing. Closer examination of employment status showed this was more than 50% missing in Turkey (with a total sample size of 11,220 and 79% missing), Mexico (with a total sample of 38,746 and 50% missing), and Comoros (n=1759 and 58% missing). For education variables, more than 50% missing data was observed in Ethiopia (n = 4,938; 58% missing), Mali (n = 4,285, 84% missing), Morocco (n= 5000, 55% missing), Republic of Congo (n = 2497, 60% missing) and Senegal (n = 3226, 54% missing) datasets. Due to these missing data considerations, data from Turkey (n=11,220) was excluded because of incomplete data and data from a further 56,444 respondents, across the globe, were excluded for missing data on other relevant variables. The final sample consisted of 207,608 individuals across 68 countries. Comparisons on the self-completion sleep question between those included in the final sample and those excluded, on employment and education, revealed similar responses suggesting that the models are generally robust to Missing At Random (see appendix 1).

Analytical strategy

Following the methodology of King et al [19] Compound Hierarchical Ordered Probit (CHOPIT) models were employed to analyse the data. The CHOPIT model incorporates the vignette and self-assessment items, allowing thresholds of the response categories from the anchoring vignettes to vary as a function of individual characteristics[29]. There are two assumptions within the vignette component, which captures information on systematic reporting bias – first, that respondents perceive the level of sleep quality represented in the vignettes in the same way; second, that there is response consistency between the vignettes and the self-assessment question ultimately asked of the respondents [29]. The self-assessment component uses information on systematic reporting bias identified from the vignette component to adjust responses. Following Tampubolon [33], a multilevel approach was adopted to account for the fact that data came from multiple countries, with a random effect included in the CHOPIT model to account for variations across countries. Models were estimated in Stata using the gllamm package and procedures outlined in Rabe-Hesketh and Skrondal [34]. The models also controlled for self-rated health and number of comorbidities.

Results

As can be seen in Table 1, the final analytic sample consisted of 207,608 individuals of which 109,332 were male (52.7%). The majority were married/cohabiting (65.8%) and employed (60.8%).

[Insert table 1 here]

<u>Unadjusted models</u>

Consistent with previous literature, gender, age, marital status and health were all significantly associated with sleep quality judgements in unadjusted models. Those who were unemployed reported poorer sleep quality (B -0.058, s.e. 0.006) than those employed. Those who were married/cohabiting reported better sleep than those who were single (B -0.034; s.e. 0.008); whereas those who were widowed (B 0.062; s.e. 0.013) and separated/divorced (B 0.051; s.e. 0.013) reported poorer sleep relative to those who were single.

[Insert Table 2]

Patterns also appeared relatively stable across countries, with the majority of respondents across all countries reporting no problems sleeping (figure 1). Maps of severity (figure 2) and formal multilevel analysis also suggest limited country differences; with approximately 3.5% of the variation due to countries in unadjusted models.

[Insert Figure 1 and Figure 2]

Vignette models

In terms of reporting heterogeneity the results were mixed. Adjusted estimates were generally similar to the unadjusted estimates (table 2). Those who were unemployed reported poorer sleep quality (B -0.092, s.e. 0.008) than those employed. Those who were married/cohabiting still reported better sleep than those who were single (B -0.058; s.e. 0.010) and those who were widowed (B 0.044, s.e. 0.018) and separated/divorced (B 0.064 s.e. 0.018) still reported poorer sleep relative to those who were single. Patterns of response to the vignettes also largely appeared similar across countries (see appendix 3). This suggests consistency in how people understand, and report sleep quality and that there is comparability across individuals, places and spaces.

Both stable and variable

Whilst much of the vignette analysis suggested that sleep quality judgements cohere, there were indications of group differences in how the concept operates. Although effect sizes were small, the results suggest that the association between relationship status and sleep is (in part) due to a difference in the way that good/poor sleep is considered. Findings for those who were widowed, for example, changed from B 0.062 (s.e. 0.013***) to B 0.044 (s.e. 0.018*) when ratings were adjusted by the anchoring vignettes. However, the most notable finding was with education. The link between education and self-reported sleep quality reduced from B -0.005 (s.e. 0.001***) to B -0.000 (s.e. 0.001) following adjustment for reporting heterogeneity (Table 2). Estimates of threshold parameters (appendix 2) suggest those with more years of education had a slightly increased threshold for reporting mild sleep problems (B 0.005; s.e. 0.001) but a lower threshold for reporting sleep problems as extreme (B -0.007; s.e. 0.001). Whilst there was limited country variability, the proportion of variability between countries also reduced in the adjusted models. This suggests some differences across countries in the ways in which people consider the concept of sleep quality. This is most notable in vignette two where there is greater variation in distributions (see appendix 3).

Discussion

This paper aimed to examine, ‘what is sleep quality’ and ‘does it operate in different ways depending on social location, context and function’? This has important implications. For one, if social position influences how we conceive of sleep problems then known inequalities in sleep will – in part – be an artefact of this.

The findings remained similar across unadjusted and adjusted models suggesting that there are inequalities in sleep quality judgements across age, gender, employment and marital status. Findings also provided support for Tareque et al’s claim that the survey questions are valid. Tareque et al[29] analysed data from Japan and found subtle changes between adjusted and unadjusted models “Before correcting for cut-point shifts using vignettes, older age, women, and the presence of comorbidities were significantly associated with a greater severity of sleep-related problems.

After correction, the associations with age and the presence of comorbidities remained significant; however, the association with gender was no longer significant." Taraque et al [29] considered these findings somewhat inconsequential and concluded that there was no need to adjust for heterogeneity in sleep studies.

At one and the same time, the current findings also add complexity to these points. Returning to the questions of 'what is sleep quality' and 'does it operate in different ways for different people', it would appear that it sits somewhere between complete coherence and complete heterogeneity. Substantive changes were detected across different models – most notably with years in formal education. Research has consistently found that those with higher education have significantly higher sleep efficiency [35] and Arber et al [24] have suggested that this may be related to increased awareness of strategies that can be used to improve sleep. The more educated may be more proactive in attempts to enhance sleep quality as well as having greater recognition of the importance of sleep for health and well-being. Within the current study, the relationship between education and sleep was attenuated once models adjusted for vignette responses. This does suggest that a key part of the pathway relates to perceptions of 'good' and 'bad' sleep; however, this is not a simple linear relationship between higher education and higher (or lower) thresholds.

Sleep quality judgements are therefore 'liminal' and neither completely stable nor completely heterogenous. This has two immediate, interrelated, implications: first, it suggests a need to move away from one dimensional ideas about the determinants of sleep/quality. Socio-ecological models of sleep recognise that sleep health is, in part, determined by societal factors (such as technology, racism, economics and the natural environment); as well as social factors (such as home, relationships, networks and socio-economic status) and individual factors (such as genetics, beliefs and attitudes)[26]. The current findings suggest that thresholds for good and poor sleep are – in part – a product of each level separately and an entanglement of all three. Several types of mechanism may be important; including *inter alia* differences and similarities in meanings attributed to sleep health (epistemological), complexities surrounding how we should sleep (normative) and the ontological issue of definition and what counts as ‘sleep’ within and between countries. More research is needed to determine how and why sleep quality judgements are able to remain consistent in some instances and not others.

A second point relates to implications for policy and interventions. In their discussion of public mental health, Drake and Whitely [36] suggest that ground-up approaches may better serve the goals of public health. Such approaches are developed by local stakeholders and communities and prioritise local knowledge, competence and resources. The current findings confirm that something similar is needed within the arena of sleep health. Rather than implementing broad scale interventions in coordination with key stakeholders [37] interventions need to be co-created with these local communities and stakeholders from the beginning. The results also suggest that vignettes would provide an appropriate methodological tool for facilitating this[i]. Whilst researcher created vignettes can be embedded within surveys to capture heterogeneity, the process of co-creating vignettes with participants can also shed light on shifting thresholds and the values which are shared within and between communities.

Limitations

The main limitations in the present study are that the data was historical and the vignette may not be considered a standard sleep question. Whilst it is important to recognise this, the aim was not to report on the current state of sleep, rather to explore a conceptual question – and to ask whether sleep quality is a concept that coheres across people, space and place. This data offers an important step forward in this initiative.

Conclusions

Overall, the results suggest a complex relationship between socio-demographic factors and self-reported sleep quality. Age, gender, employment status, marital status and level of education all impact on subjective sleep quality judgements – but, for some factors, this is partly because they influence the thresholds which separate good and bad sleep. Notwithstanding the need for further research, this adds further weight to the idea that if we are to develop public health responses to sleep we need to do this from the ground up [38].

Acknowledgments

This article used data from the World Health Organisation.

References

1. Ramlee, F., A.N. Sanborn, and N.K. Tang, *What sways people's judgment of sleep quality? a quantitative choice-making study with good and poor sleepers.* Sleep, 2017. **40**(7).
2. Harvey, A.G., et al., *The subjective meaning of sleep quality: a comparison of individuals with and without insomnia.* Sleep, 2008. **31**(3): p. 383-393.
3. Clark, A.J., et al., *Onset of impaired sleep and cardiovascular disease risk factors: a longitudinal study.* Sleep, 2016. **39**(9): p. 1709-1718.
4. Bin, Y.S., *Is sleep quality more important than sleep duration for public health?* Sleep, 2016. **39**(9): p. 1629-1630.
5. Buysse, D.J., et al., *The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research.* Psychiatry research, 1989. **28**(2): p. 193-213.
6. Åkerstedt, T., et al., *The subjective meaning of good sleep, an intraindividual approach using the Karolinska Sleep Diary.* Perceptual and motor skills, 1994. **79**(1): p. 287-296.
7. Ohayon, M., et al., *National Sleep Foundation's sleep quality recommendations: first report.* Sleep health, 2017. **3**(1): p. 6-19.
8. Ohayon, M.M., et al., *A provisional tool for the measurement of sleep satisfaction.* Sleep health, 2018. **4**(1): p. 6-12.
9. Ogeil, R.P., et al., *Assembling a 'good'and 'bad'night's sleep: A multifactorial proposition.* Lifestyle Medicine: p. e48.
10. Nettleton, S., *The sociology of health and illness*. 2020: John Wiley & Sons.
11. Stempsey, W.E., *Emerging medical technologies and emerging conceptions of health.* Theoretical Medicine and Bioethics, 2006. **27**(3): p. 227-243.
12. Blaxter, M., *Health & Lifestyles Routledge*. 1990, London.
13. Landry, G.J., J.R. Best, and T. Liu-Ambrose, *Measuring sleep quality in older adults: a comparison using subjective and objective methods.* Frontiers in aging neuroscience, 2015. **7**: p. 166.
14. Svetnik, V., et al., *How well can a large number of polysomnography sleep measures predict subjective sleep quality in insomnia patients?* Sleep medicine, 2020. **67**: p. 137-146.
15. Meadows, R., *The 'negotiated night': an embodied conceptual framework for the sociological study of sleep.* The Sociological Review, 2005. **53**(2): p. 240-254.
16. Coveney, C., S.J. Williams, and J. Gabe, *Medicalisation, pharmaceuticalisation, or both? Exploring the medical management of sleeplessness as insomnia.* Sociology of health & illness, 2019. **41**(2): p. 266-284.
17. Star, S.L. and J.R. Griesemer, *Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39.* Social studies of science, 1989. **19**(3): p. 387-420.
18. Ebeling, M.F., *Mediating uncertainty: Communicating the financial risks of nanotechnologies.* Science Communication, 2008. **29**(3): p. 335-361.
19. Trompette, P. and D. Vinck, *Revisiting the notion of boundary object.* Revue d'anthropologie des connaissances, 2009. **3**(1): p. 3-25.
20. Riemann, D., et al., *European guideline for the diagnosis and treatment of insomnia.* Journal of sleep research, 2017. **26**(6): p. 675-700.
21. Voderholzer, U., et al., *Are there gender differences in objective and subjective sleep measures? A study of insomniacs and healthy controls.* Depression and anxiety, 2003. **17**(3): p. 162-172.
22. van den Berg, J.F., et al., *Sex differences in subjective and actigraphic sleep measures: a population-based study of elderly persons.* Sleep, 2009. **32**(10): p. 1367-1375.

23. Baker, F.C., S. Maloney, and H.S. Driver, *A comparison of subjective estimates of sleep with objective polysomnographic data in healthy men and women.* Journal of psychosomatic research, 1999. **47**(4): p. 335-341.
24. Arber, S., M. Bote, and R. Meadows, *Gender and socio-economic patterning of self-reported sleep problems in Britain.* Social science & medicine, 2009. **68**(2): p. 281-289.
25. Fatima, Y., et al., *Exploring gender difference in sleep quality of young adults: findings from a large population study.* Clinical medicine & research, 2016. **14**(3-4): p. 138-144.
26. Grandner, M.A., *Sleep, health, and society.* Sleep medicine clinics, 2017. **12**(1): p. 1-22.
27. Taylor, B., *Unconsciousness and society: The sociology of sleep.* International Journal of Politics, Culture, and Society, 1993. **6**(3): p. 463-471.
28. Huang, Z. and F. Kämpfen, *The association between depressive symptoms and self-reported sleep difficulties among college students: Truth or reporting bias?* PloS one, 2021. **16**(2): p. e0246370.
29. Tareque, M.I., et al., *Is adjustment for reporting heterogeneity necessary in sleep disorders? Results from the Japanese World Health Survey.* BMC psychiatry, 2016. **16**(1): p. 1-7.
30. Chevalier, A. and A. Fielding, *An introduction to anchoring vignettes.* Journal of the Royal Statistical Society: Series A (Statistics in Society), 2011. **174**(3): p. 569-574.
31. King, G., et al., *Enhancing the validity and cross-cultural comparability of measurement in survey research.* American political science review, 2004. **98**(1): p. 191-207.
32. Subramanian, S.V., T. Huijts, and M. Avendano, *Self-reported health assessments in the 2002 World Health Survey: how do they correlate with education?* Bulletin of the World Health Organization, 2010. **88**: p. 131-138.
33. Tampubolon, G., *Multilevel anchoring vignettes for comparative study of health inequalities in 48 countries of the WHO.".* 2010.
34. Rabe-Hesketh, S. and A. Skrondal, *Estimating CHOPIT models in GLLAMM: Political efficacy example from King et al.(2002).* London: Institute of Psychiatry, King's College, 2002.
35. Sosso, F.A.E., S.D. Holmes, and A.A. Weinstein, *Influence of socioeconomic status on objective sleep measurement: A systematic review and meta-analysis of actigraphy studies.* Sleep Health, 2021. **7**(4): p. 417-428.
36. Drake, R.E. and R. Whitley, *Building behavioral health systems from the ground up.* World Psychiatry, 2015. **14**(1): p. 50.
37. Hale, L., W. Troxel, and D.J. Buysse, *Sleep health: An opportunity for public health to address health equity.* Annual review of public health, 2020. **41**: p. 81-99.
38. Meadows, R., et al., *Counting sleep? Critical reflections on a UK national sleep strategy.* Critical Public Health, 2021. **31**(4): p. 494-499.

Figure 1: Self Rated Sleep reports for each country

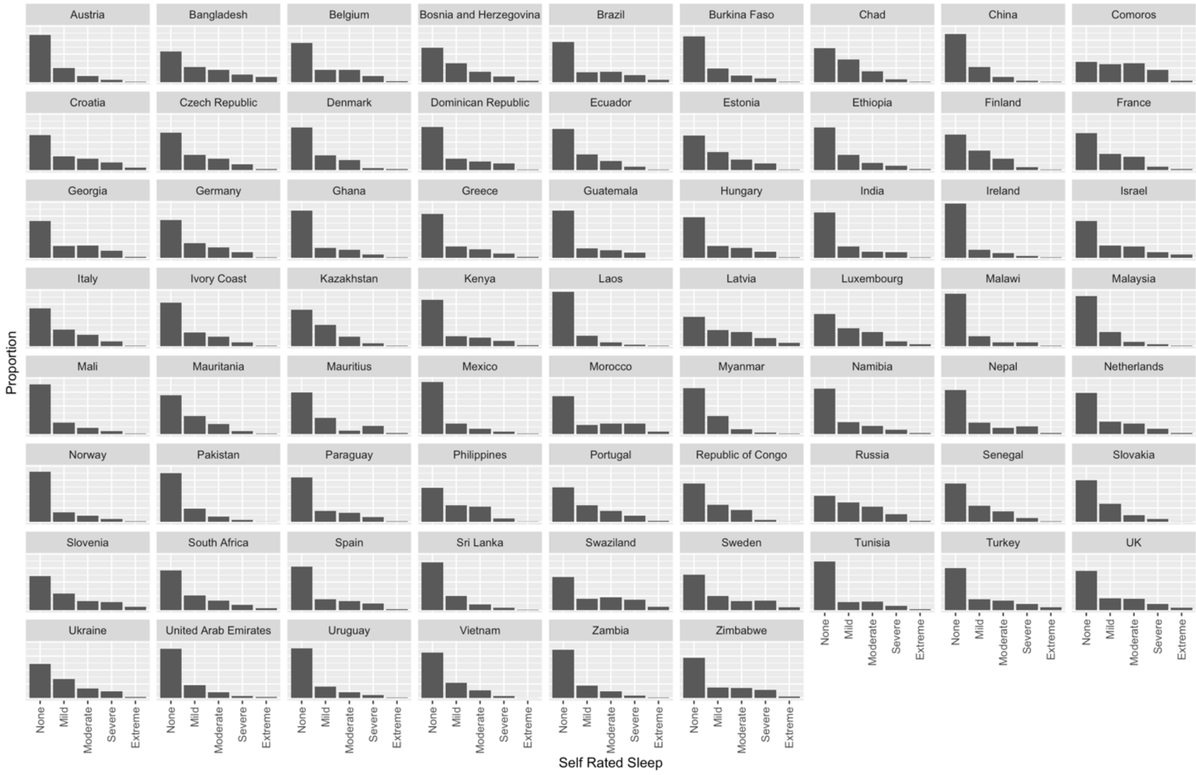


Figure 2: The percentage of people in each country reporting moderate, severe or extreme problems.

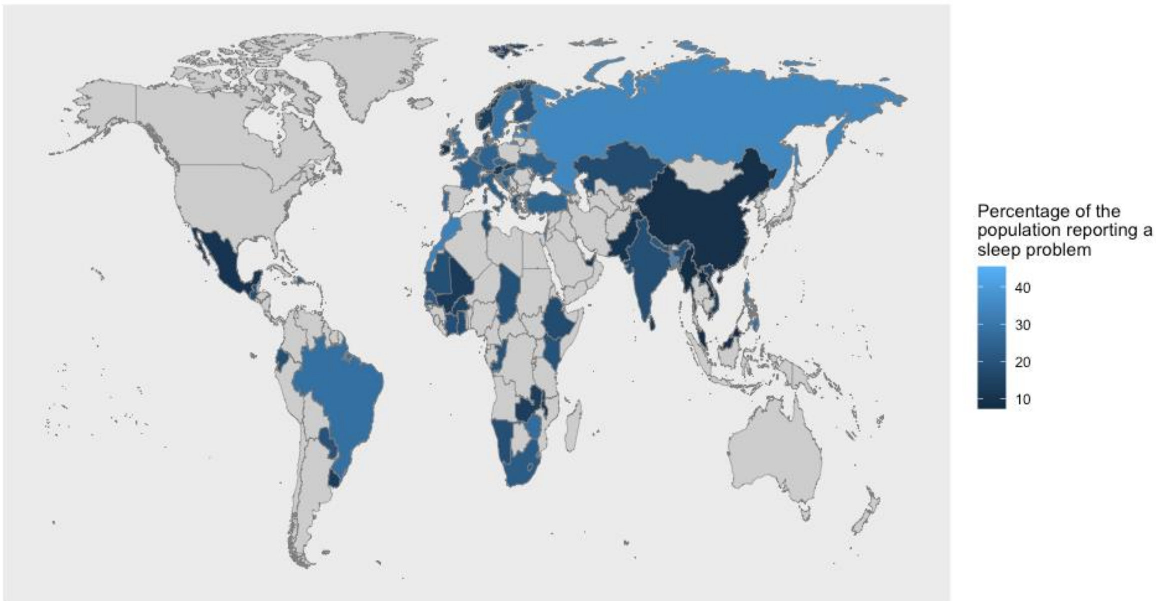


Table 1 – Sample characteristics

	Total (N=207608)
Gender[1]	
Male	109332 (52.7%)
Female	98276 (47.3%)
Age in years[2]	40.68 (16.25): 15.00-101.00
Marital status	
Single	41846 (20.2%)
Married/Cohabiting	136564 (65.8%)
Separated/Divorced	11849 (5.7%)
Widowed	17349 (8.4%)
Employment status	
Employed	126227 (60.8%)
Unemployed	81381 (39.2%)
Education (years)	7.65 (5.27): 0.00-50.00
Self-rated health (higher is worse)[3]	2.23 (0.91): 1.00-5.00
Number of comorbidities[4]	1.10 (1.39): 0.00-7.00

[1] Categorical summaries report N (%)

[2] Continuous summaries report Mean (SD): Range

[3] Question wording: In general, how would you rate your health today? (1 = very good, 2 = good, 3 = moderate, 4 = bad, 5 = very bad)

[4] Number of morbidities experienced – back pain; angina pectoris; pain or discomfort in chest; asthma; wheezing or whistling breathing; sadness, emptiness of depression; lack of interest

Table 2 – Unadjusted and adjusted results

	Model 1 - no vignette adjustment				Model 2 - vignette adjustment		
	B	**S.E**	**Significance[1]**		**B**	**S.E**	**Significance**
Male	-0.103	0.006	0.000		-0.093	0.008	0.000
Age	0.010	0.000	0.000		0.009	0.000	0.000
Employed	-0.058	0.006	0.000		-0.092	0.008	0.000
Married/Cohabiting	-0.034	0.008	0.000		-0.058	0.010	0.000
Separated/Divorced	0.051	0.013	0.000		0.064	0.018	0.000
Widowed	0.062	0.013	0.000		0.044	0.018	0.012
Education (years)	-0.005	0.001	0.000		0.000	0.001	0.538
Self-rated health (higher is worse)[2]	0.333	0.003	0.000		0.336	0.004	0.000
Number of co-morbidities[3]	0.221	0.002	0.000		0.208	0.003	0.000
Thresholds[4]				*Vignettes*			
i1	1.202	0.026		Vignette 1	0.636	0.016	
i2	1.907	0.026		Vignette 2	2.128	0.015	
i3	2.623	0.026		Vignette 3	2.456	0.016	
i4	3.680	0.028		Vignette 4	2.654	0.016	
				Vignette 5	3.137	0.017	
Random effects							
Log SD (vignette)					-0.002	0.004	
Country	0.035	0.006			0.020	0.002	
N (individual)	207,608				207,608		
N (Country)	68				68		

[1] Significance calculated using Wald tests (B/SE)

[2] Question wording: In general, how would you rate your health today? (1 = very good, 2 = good, 3 = moderate, 4 = bad, 5 = very bad)

[3] Number of morbidities experienced – back pain; angina pectoris; pain or discomfort in chest; asthma; wheezing or whistling breathing; sadness, emptiness of depression; lack of interest

[4] For full details of the threshold equations for model 2 see appendix 2

Appendix 1: Missing data analysis – self rated problems sleeping

	Missing employment			Missing education	
	FALSE	TRUE		FALSE	TRUE
None	62.5	63.5		63.2	57.6
Mild	18.8	17.0		18.4	19.8
Moderate	11.5	11.8		11.3	13.3
Severe	6.2	6.3		6.0	7.8
Extreme	1.0	1.5		1.0	1.6

Appendix 2: Vignette thresholds

	B	**S.E**	**Significance[1]**
Threshold i1			
Male	0.020	0.006	0.002
Age	-0.002	0.000	0.000
Employed	-0.026	0.007	0.000
Married/Cohabiting	-0.013	0.008	0.103
Separated/Divorced	0.043	0.014	0.002
Widowed	-0.016	0.014	0.276
Education (years)	0.005	0.001	0.000
Self-rated health (higher is worse)[2]	-0.003	0.004	0.368
Number of co-morbidities[3]	-0.005	0.002	0.017
Constant	1.246	0.014	0.000
Threshold i2			
Male	-0.014	0.007	0.054
Age	0.001	0.000	0.001
Employed	0.037	0.007	0.000
Married/Cohabiting	0.026	0.009	0.006
Separated/Divorced	-0.017	0.017	0.313
Widowed	0.021	0.016	0.193
Education (years)	-0.007	0.001	0.000
Self-rated health (higher is worse)[2]	-0.015	0.004	0.000
Number of co-morbidities[3]	-0.005	0.003	0.055
Constant	-0.450	0.013	0.000
Threshold i3			
Male	0.001	0.007	0.893
Age	0.002	0.000	0.000
Employed	0.009	0.007	0.174
Married/Cohabiting	-0.007	0.009	0.437
Separated/Divorced	-0.045	0.015	0.003
Widowed	-0.017	0.015	0.248
Education (years)	0.000	0.001	0.804
Self-rated health (higher is worse)[2]	0.010	0.004	0.005
Number of co-morbidities[3]	-0.008	0.002	0.001
Constant	-0.310	0.013	0.000
Threshold i4			
Male	-0.016	0.006	0.013
Age	0.000	0.000	0.159
Employed	-0.004	0.006	0.577
Married/Cohabiting	0.039	0.008	0.000
Separated/Divorced	0.005	0.014	0.717
Widowed	0.058	0.014	0.000
Education (years)	-0.007	0.001	0.000
Self-rated health (higher is worse)[2]	0.004	0.003	0.260

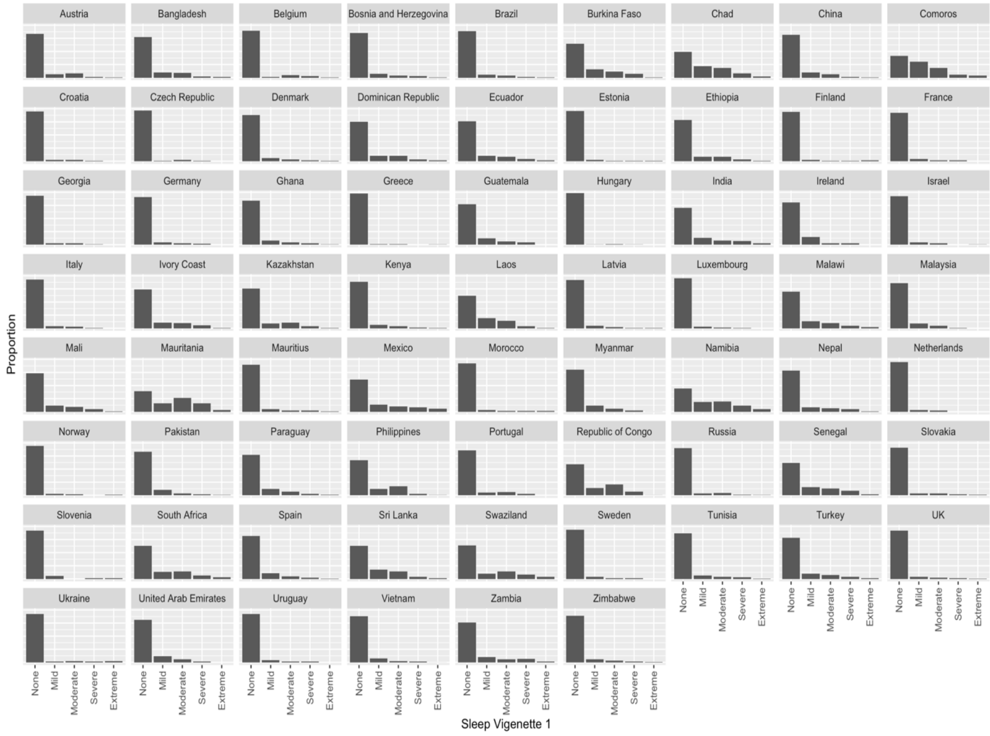
Number of co-morbidities[3]	-0.014	0.002	0.000
Constant	0.228	0.012	0.000

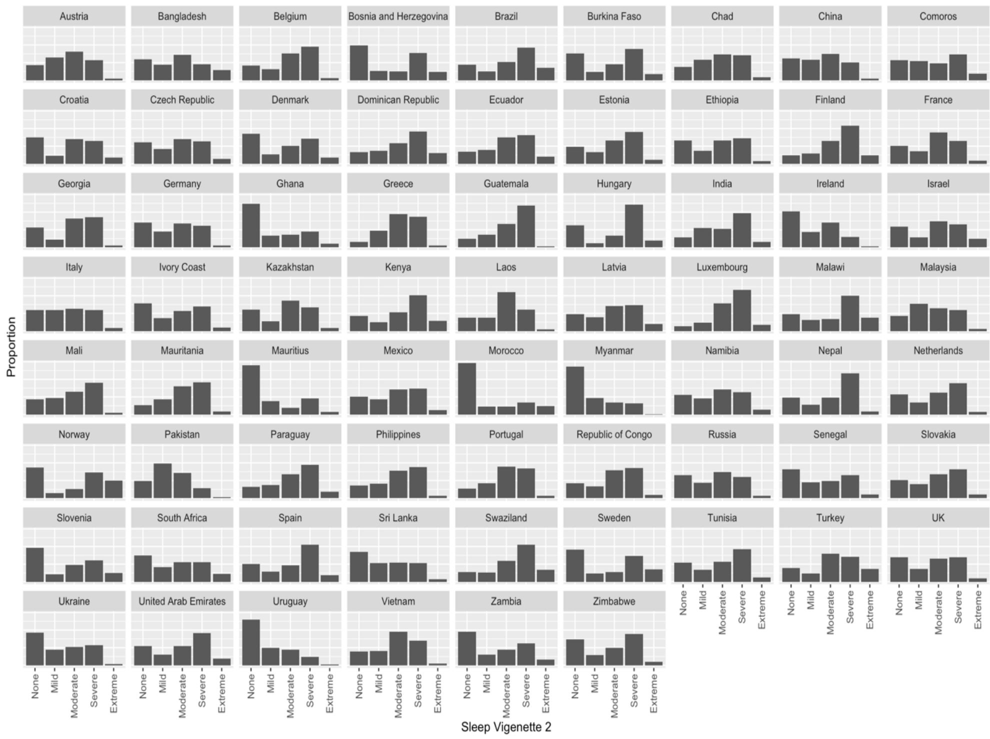
[1] Significance calculated using Wald tests (B/SE)

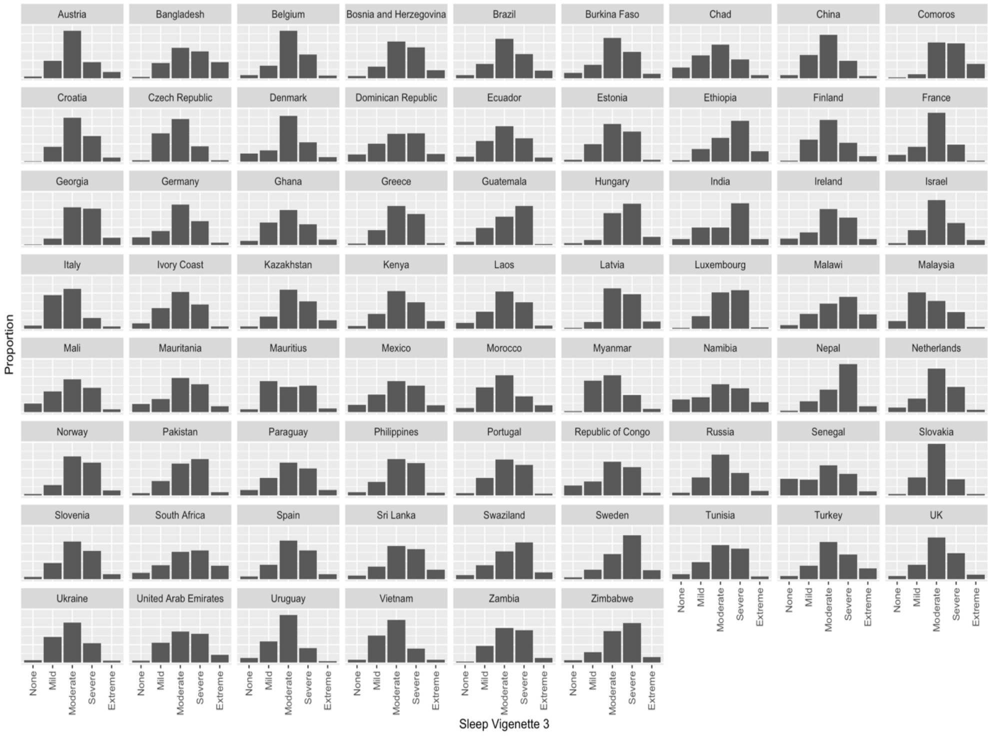
[2] Question wording: In general, how would you rate your health today? (1 = very good, 2 = good, 3 = moderate, 4 = bad, 5 = very bad)

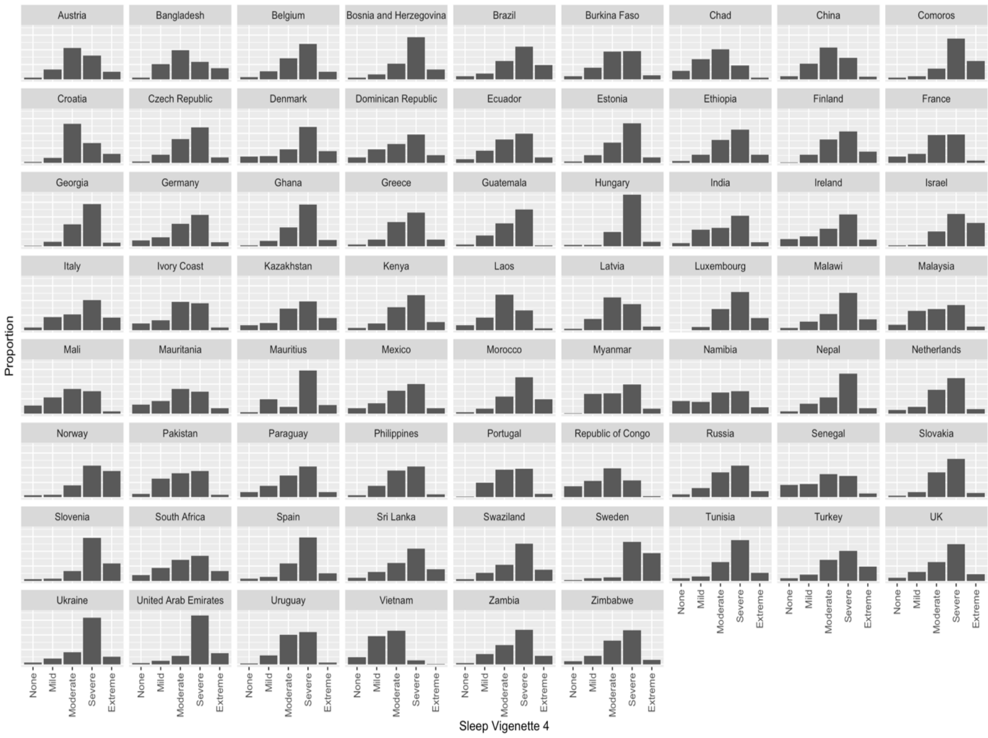
[3] Number of morbidities experienced – back pain; angina pectoris; pain or discomfort in chest; asthma; wheezing or whistling breathing; sadness, emptiness of depression; lack of interest

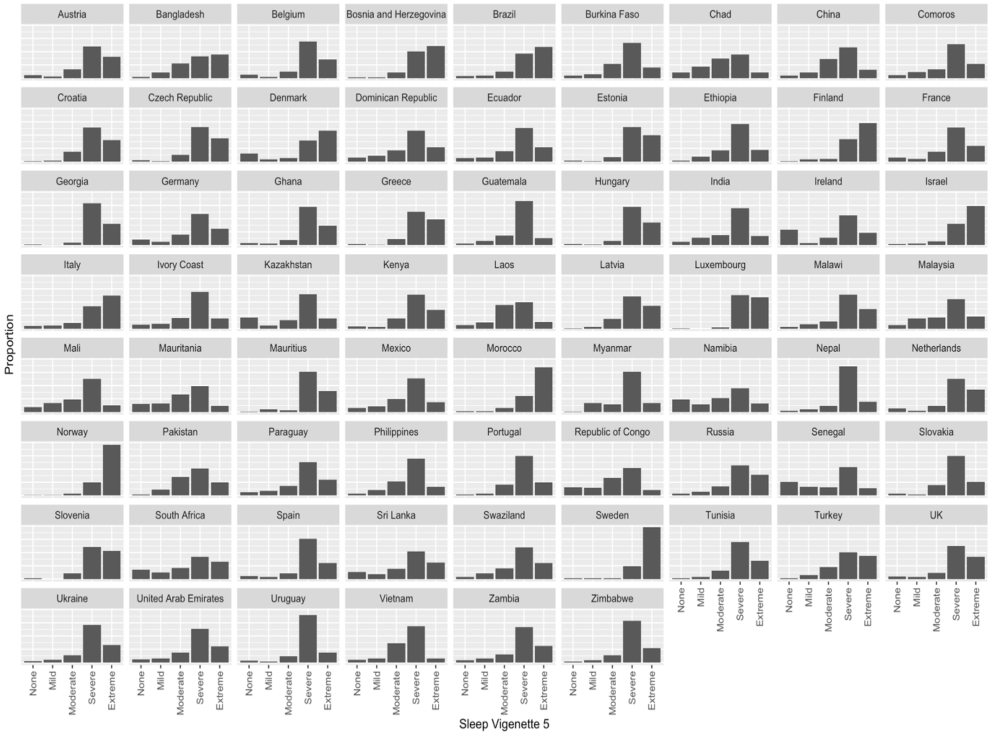
Appendix 3: Patterns of response to vignette by country






Austria
Bangladesh
Belgium
Bosnia and Herzegovina
Brazil
Burkina Faso
Chad
China
Comoros
Croatia
Czech Republic
Denmark
Dominican Republic
Ecuador
Estonia
Ethiopia
Finland
France
Georgia
Germany
Ghana
Greece
Guatemala
Hungary
India
Ireland
Israel
Italy
Ivory Coast
Kazakhstan
Kenya
Laos
Latvia
Luxembourg
Malawi
Malaysia
Mali
Mauritania
Mauritius
Mexico
Morocco
Myanmar
Namibia
Nepal
Netherlands
Norway
Pakistan
Paraguay
Philippines
Portugal
Republic of Congo
Russia
Senegal
Slovakia
Slovenia
South Africa
Spain
Sri Lanka
Swaziland
Sweden
Tunisia
Turkey
UK
Ukraine
United Arab Emirates
Uruguay
Vietnam
Zambia
Zimbabwe
Proportion
None
Mild
Moderate
Severe
Extreme
Sleep Vigenette 3


Austria
Bangladesh
Belgium
Bosnia and Herzegovina
Brazil
Burkina Faso
Chad
China
Comoros
Croatia
Czech Republic
Denmark
Dominican Republic
Ecuador
Estonia
Ethiopia
Finland
France
Georgia
Germany
Ghana
Greece
Guatemala
Hungary
India
Ireland
Israel
Italy
Ivory Coast
Kazakhstan
Kenya
Laos
Latvia
Luxembourg
Malawi
Malaysia
Mali
Mauritania
Mauritius
Mexico
Morocco
Myanmar
Namibia
Nepal
Netherlands
Norway
Pakistan
Paraguay
Philippines
Portugal
Republic of Congo
Russia
Senegal
Slovakia
Slovenia
South Africa
Spain
Sri Lanka
Swaziland
Sweden
Tunisia
Turkey
UK
Ukraine
United Arab Emirates
Uruguay
Vietnam
Zambia
Zimbabwe
Proportion
None
Mild
Moderate
Severe
Extreme
Sleep Vigenette 4



[i] We would like to thank one of the reviewers for helping us bring this point out more fully