
Sleep Deprivation: Practical and Philosophical Considerations

1

Matt T. Bianchi

Introduction

Despite being a fairly young field, sleep medicine has made enormous progress from mechanistic to applied clinical sciences. In this volume, the literature linking sleep to a diversity of health and performance topics is explored. The growth and development of this field has been explored in several books, and the interested reader is directed to explore these general readership works that nevertheless capture the evolution of sleep science and its relation to medicine [1, 2]. As a complement to the numerous textbooks of sleep medicine, these accounts provide an important historical perspective. Such context is particularly interesting because sleep may be unique among medical subspecialties in that it has a nearly universal audience in the lay-community, and knowledge about sleep is claimed as much from personal or cultural experience as it is from careful experimentation. This is both a challenge and an opportunity at the intersection of academic research, clinical practice, and social behavior. It is telling that in the annual meeting of the Associated Professional Sleep Societies in 2013, there was a symposium dedicated to the history and science of segmented sleep and the (arguably mythical) assumption that sleep should be (or at

least feel like it is) uninterrupted. Among the speakers was historian A. Roger Ekirch, author of “At Day’s Close: Night in Times Past” [3], who provided intriguing context to the presentations by leaders in the field. Although the topics presented in this volume focus on the scientific and medical perspectives, the clear relevance for wellness and performance has broad relevance beyond these arenas.

The expanding knowledge base in this field may enjoy more rapid dissemination precisely because of the universality of sleep itself. The narratives emerging from new research, particularly in the area of sleep deprivation, carry immense personal valence and strong apparent face validity. This has positive and negative consequences: information dissemination may have fewer hurdles in the way of relevance and believability, but the risk of bias in the narrative may be more difficult to mitigate. It is not hard to imagine lay-targeted headlines that would easily capture unchallenged attention, like “No one wants a sleepy surgeon,” or “Everyone knows how badly it feels to be sleep deprived.” Even such apparently “obvious” narratives have alternative or competing narratives that may also, in isolation, seem quite compelling. Consider the hypothetical headline, “Patients prefer professionally dressed physicians”—it may seem like an obvious finding, especially for patients forced to choose between professional versus casual attire. But what if the question asked if you prefer a professionally dressed physician or an empathetic one? Taking the query one step further: how _

M.T. Bianchi, M.D., Ph.D., M.M.Sc. (✉)
Division of Sleep Medicine, Department of Neurology,
Massachusetts General Hospital, Boston, MA, USA
e-mail: mtbianchi@partners.org

well-dressed would a physician have to be to make up for lack of empathy, or how empathetic would a physician have to be to make up for casual attire? Now reconsider the sleepy surgeon: what if the choice were between sleep deprived yet invested in your care and thoroughly familiar with your case, and a night shift “covering” physician who is neither invested nor familiar with your case. This issue has been raised regarding queries about sleep, by considering the difference between asking whether you would like more sleep, versus asking what waking activities would you give up for more sleep [4]. Placing sleep in a broader context through a trade-off approach revealed in a cleverly designed survey study that few people chose sleep over potentially attractive alternative activities, despite a high prevalence of apparent sleep complaints [5]. The exercise of considering a broader context, including risk-benefit trade-offs, strengthens the narratives and insulates against the insidious risk of overcommitting to a particular narrative. Modern medicine has recognized that face validity and personal experience have a worthy competitor for our attention in the form of careful experimentation. The history of medicine is littered with examples of expert consensus later exposed as folly when carefully studied. Even the findings of well-intended clinical trials in the modern era are often not replicated [6], so from a Bayesian standpoint we might do well to collectively approach the biomedical literature with skepticism because the prior probability seems to favor refutation rather than confirmation.

With the goal of cautious optimism, this chapter outlines some key ideas to keep in mind as one explores the remaining contents. A series of recent debates and editorials capture the sobering reality that studying the role of sleep in health and disease is no simple undertaking. The interested reader is encouraged to sample these engaging discourses directly, concerning the importance of sleep in general [7, 8], the challenges in studying short sleep [9], the concept of sleep debt versus adaptive regulation [4], and the possibility of enhancing health by improving sleep [10].

What Is Normal Sleep?

Identifying what is normal sleep is not as simple as one might hope, yet it is the foundation of any discussion of sleep deprivation. What is considered normal may evolve over time as research findings help disentangle what is “common” from what may be associated with adverse health outcomes. Consider blood pressure, blood sugar, and cholesterol—these are examples of continuous variables with evolving thresholds partitioning health and disease. Likewise, the many facets of sleep physiology may be best understood as a distribution of values, the tails of which represent (perhaps blurry) transitions to disease status.

It has been suggested that presence of symptoms plays an important role in defining pathology, whether in the historical use of “syndrome” suffix for the metrics of obstructive sleep apnea (OSA) combined with sleepiness (which is no longer required [11]), or in more recent discussions about short sleep duration [10]. The symptom-focused approach seems sensible at first glance, but on closer consideration we face problems of inference. If sleepiness is that which occurs when we lack sleep and resolves when we get sleep, then if there is no sleepiness, there is no sleep problem. Yet if we don’t accept that lack of sleep-related symptoms implies normal sleep in some settings (e.g., OSA), we should be cautious implying that short (or long) sleepers should be subdivided into normal versus abnormal based solely on symptoms. The commonly used Epworth Sleepiness Scale has minimal correlation with objective measures of sleep [12, 13]. Also, it has been shown that many patients with even severe OSA do not report sleepiness [14, 15], yet untreated OSA harbors adverse health risks regardless [16]. It may be that symptoms are by definition required for certain disorders (such as restless legs or insomnia), and they may also help phenotype individuals, perhaps based on vulnerability to challenges such as sleep restriction or OSA. However, some sleep disorders may be asymptomatic, and some symptoms (such as sleepiness or fatigue) are not

specific to sleep problems. The medical canon is filled with examples of asymptomatic phases of chronic diseases, and certainly the field of sleep disorders is no exception.

If one could state what the normal quantity and quality of sleep was, the discourse would surely include a caveat that the answer might vary among individuals, not only in the sense that a distribution of values might be acceptable, but also in the sense that a given value might be normal for only some people or only in some settings. For some individuals, 6 h of sleep per night could be normal, while for others, restriction to 6 h per night would incur substantial symptoms; the former group might even feel worse with the extra 2 h of sleep. Does the body care about total sleep time, or the sleep stage content, or continuity? Does stage content only matter when TST is restricted? Are different organ systems, or different brain functions, differentially sensitive? Would the answers to these questions change between individuals, or even within an individual depending on health status, recent sleep history, or the consumption of caffeine or alcohol? The combinatorial possibilities are daunting. Thus, defining normal sleep, whether by total duration, stage content, arousals, breathing, or other metrics, is not a trivial question.

Assuming the “basic” question of what is normal sleep can be answered, one must then identify how much deviation from normal is relevant? The issue of defining relevance can also be considered as a spectrum, ranging from that which is noticeable but either tolerable or overcome with simple countermeasures, to that which impairs performance, and eventually that which tangibly compromises medical or psychiatric health. One would like to know whether deviations from normal are sensed by the body in an absolute manner (say, one less hour of sleep), or in a relative manner (say, 10 % less sleep)?

The Act of Measuring Disturbs the System Under Observation

The experimental literature on the performance impact of sleep deprivation may be influenced by the Hawthorne effect, in which subjects may behave or perform differently when under obser-

vation. There may also be factors that reduce performance in experimental settings, such as lack of interest, tedium of the task at hand, and so forth. The extent to which this may play a role in extrapolating experimental results to real-world situations, especially when effect sizes are small, remains open to debate.

However, there is an even more fundamental issue at stake when we record sleep using PSG, as outlined in a recent article analogizing this gold standard test with quantum uncertainty [17]. It is obvious to many patients experiencing the sleep laboratory, regardless of their background physics training, that observing the sleep-wake system perturbs it in proportion to the burden and invasiveness of the measurement tools. One well-known example is the so-called “first night effect,” in which the laboratory environment tends to increase N1, decrease sleep efficiency, and decrease REM sleep. However, it is also worth noting that some patients with insomnia may exhibit a “reverse” form of this, in which their sleep is actually improved in the laboratory setting despite the unusual environment. This is presumed to occur because one or more factors contributing to insomnia in the home setting are not present in the laboratory [18]. The recurrent theme of trade-offs thus surfaces both clinically and experimentally in the very question of how we measure sleep.

Sleep Debt, Sleep Extension, and Sleep Restriction

The topic of sleep debt raises interesting questions about the experimental investigation of sleep loss. Observations of sleep duration extension when provided the opportunity of extra time in bed have been interpreted to imply baseline sleep debt. That narrative assumes that the body precisely regulates the amount of sleep it needs, without capacity to adapt. In other words, more sleep cannot occur, even if the circumstances allow, without sleep debt. This logic hardly holds in other domains, such as hunger and food intake compared to caloric needs, as elegantly argued by Horne in his recent discourses suggesting that sleep duration is adaptive and depends on context and waking needs [4, 19].

Sleep extension beyond the acute setting may be feasible in small amounts (perhaps 1–2 h), but when the total time in bed exceeds physiological sleep capacity, fragmentation and decreased sleep efficiency ensue [20]. This should come as no surprise, as the technique of sleep restriction therapy is aimed at reversing the self-reinforcing trend among some insomniacs who make the mistake of spending more time in bed than their sleep capacity, thus perpetuating the pattern of initiation and/or maintenance sleep difficulties.

Numerous studies have investigated the impact of multiple nights of sleep restriction on physiological and performance outcomes, many of which are described in the chapters of this volume. Although the studies differ in methodology, nearly all of them report impairments, with one of the most commonly cited studies suggesting that even minor (6 h per night) restriction results in accumulated sleep debt equivalent to total sleep deprivation [21]. That study is also commonly cited as evidence that subjective sleepiness ratings underestimate objective performance metrics. However, other literature suggests that sleep restriction via gradual reduction of sleep time, in naturalistic home environments, was not only well tolerated, but also participants actually maintained the schedule voluntarily for at least 1 year following the studies [4].

The Differential Diagnosis of Self-Reported Sleep Duration: Lumping Versus Splitting

Because self-reported sleep duration is so important clinically and epidemiologically, it is a useful exercise to consider the potential underlying phenotypes for individuals reporting short sleep duration. While lumping by sleep duration may be convenient and feasible, when one considers the differential diagnosis of sleep duration, the splitting counter-argument is compelling. The possible phenotypes lumped into a group called “short sleep” could include at least the following:

1. Accurate reporting of the average of consistent objective short sleep time
2. Accurate reporting of objective short nocturnal sleep time without taking into account naps

3. Accurate reporting of the average of highly fluctuating sleep times
4. Underestimation relative to a longer objective sleep duration due to misperception insomnia
5. Underestimation relative to a longer objective sleep duration due to errors in reporting

Any of these categories could be further split according to the presence or absence of comorbid sleep disorders such as sleep apnea. Additional splitting could incorporate comorbid medical or psychiatric pathology, medications, age, genetic variance in susceptibility to sleep deprivation, and so forth. Comorbidities could influence the impact of sleep duration on health or potentially even on the accuracy of subjective reporting of sleep duration. Many survey studies attempt to control for comorbidities, but underlying sleep disorders are difficult to assess by survey, especially the disease with arguably the most dramatic objective sleep disruption—sleep apnea. The downside to splitting is of course that the sample sizes needed to explore the combinatorial possibilities rise rapidly.

The lumping approach may alter epidemiological correlations with various outcomes. Indeed, recent data suggests that medical morbidity is mainly associated with objective short sleep duration [22]. However, even this finding requires further inquiry—given the night to night variability of sleep, and of insomnia, it could be that short sleep duration in the lab is as much a marker of sensitivity to environmental challenge (i.e., the vulnerability of sleep in general) rather than a direct link to pathology.

One can undertake a similar differential diagnosis exercise with self-reported long sleep durations. In epidemiological surveys of sleep, long durations also correlate with adverse health outcomes (although this does not often resonate with media accounts focusing on the narrative that we need more sleep as a society). There has been much speculation as to the underlying reasons for U-shaped associations [9, 23, 24]. In many cases the longer self-reported sleep durations show greater health risk than shorter durations [24], as is the case for all-cause mortality (1.1 vs. 1.23), cardiovascular mortality (1.06 vs. 1.38), and cancer mortality (0.99 vs. 1.21). Even if we assume that short and long sleep

self-reports are accurate, and that sleep duration correlates with incident adverse health outcomes, one must resist the inferential temptation to conclude that altering sleep duration will reverse or reduce these risks, which remains untested [10].

Experimental Sleep Deprivation Versus Clinical Insomnia

Patients with insomnia represent a natural target population for extrapolating the findings from experimental sleep deprivation studies. Grandner et al. recently reviewed the literature of self-reported and laboratory-measured short sleep, including an excellent overview of the challenges in this domain [9]. One particular issue regarding clinical extrapolation is that experimental sleep disturbance in a healthy individual is not analogous to the lack of sleep and hyperarousal associated with insomnia [25]. Sleep restricting a health adult generally results in objective hypersomnia (e.g., by multiple sleep latency testing), but this is not commonly observed in patients with insomnia. It is noteworthy that demonstrating objective consequences of insomnia has not enjoyed the success of demonstrating the impact of experimental deprivation. In fact, a recent review captures this challenge in its title, “Searching for the daytime impairments of primary insomnia” [26]. This is perhaps not surprising, when one considers by comparison that the dramatic physiology of severe sleep apnea, with recurrent arousals and hypoxia, does not correlate well with daytime sleepiness.

Correlation and Causation

It should go without saying that correlation is not causation, yet even modern literature sometimes offers exceptions to this sacred dictum. On one hand, there is a vast literature of carefully controlled laboratory experiments, manipulating the sleep of highly selected individuals living in highly unusual environments. In this world, we

are as close to causation as can be expected in human research. On the other hand, we have decades of epidemiological studies of self-reported sleep habits, such as napping or sleep duration. In this world, even with prospective studies, if there is no randomization then causation is nowhere to be found, no matter how large the study or how small the p -value. From either of these worlds, extrapolating the findings to the worlds of clinical practice and operational guidelines is arguably the most important challenge facing the field. The extent of control in an experimental paradigm can be taken as a good estimate of the extent to which the results will not generalize to other conditions. Consider a light pulse given during a dim light constant routine experiment, which might dramatically shift the circadian clock; the same light pulse might go completely unnoticed in the background of potentially wild light exposure fluctuations in a real-world day. A striking example of experimental-versus-naturalistic dissociation emerged from a study showing non-rhythmic activity in mice over natural outdoor light–dark cycles in which mice self-chose their light exposure [27]. This observation is in striking contrast to the imposed unnatural step-function light–dark cycles of the modern rodent lab. This does not mean that circadian rhythms are an artifact of lab conditions, but it does mean that the system is so flexible (or noisy) that the rhythms are not always robustly manifested. Of course the problem of external validity is not limited to experimental investigations—the patient undergoing clinical PSG may exhibit distinctly different physiology in the home setting, where caffeine, alcohol, or other factors may differ from that observed during clinical testing, yet clinical decisions are often based on the laboratory data.

Numerous heuristics and biases impact the subjective response to seemingly straightforward questions about sleep duration often employed in epidemiological studies of sleep, not to mention the myriad factors impacting the objective sleep duration and even whether the subjective report matches the objective duration. Sleep duration is commonly underestimated especially among

insomniacs [28], and the underestimation can be exaggerated when the time frame over which the estimate is requested is increased [29, 30]. Even if one could assume accuracy of self-reported sleep duration, duration is only one dimension of sleep and does not take into account sleep quality, sleep pathology, or individual susceptibilities to sleep disturbance or restriction. Even if simple duration is important [22], it is an entirely different question as to whether increasing sleep duration would abrogate medical risk(s). Randomizing individuals to such sleep interventions for extended periods (months or years) would not be feasible (or blind-able). Naturalistic observational studies provide an alternative perspective, but still cannot prove causality. For example, tracking patients longitudinally could reveal subsets of patients who either shorten or lengthen their sleep duration over the observation period, and this could be linked to some outcomes of interest [31]. However, without randomization, one cannot establish whether a factor that prompted a patient to “naturally” assume a course of action was itself responsible for the outcome of interest, rather than the course of action itself.

Other Statistical and Methodological Considerations

An entire text could be devoted specifically to the statistical pitfalls commonly encountered in biomedical research (and indeed they exist [32, 33]). It is worth, however, mentioning certain topics of relevance that may not be commonly discussed. The first deals with a common currency of research findings: the effect size. The magnitude of an observed effect is central to clinical research. The mistake of equating statistical significance and clinical (or philosophical) significance is so common, and has prompted so much editorializing, it seems almost trite to write about it yet again here. Clinical and statistical “significance” can dissociate under conditions when small p -values accompany miniscule effect sizes, usually in large sample studies. It is arguably better to observe a nonsignificant p -value, with a 95 % confidence interval of the effect size

that spans a clinically meaningful value, than to obtain a very small p -value for a marginal effect size with a narrow confidence interval that does not include a clinically significant value. The former leaves open the possibility of an important effect, perhaps in a future study, while the latter convincingly suggests the effect can be ignored. In 1994, Cohen himself lamented that despite decades of severe criticism, the null-hypothesis testing (p -value) strategy has not yet been eradicated [34]. We have recently added to the lamenting literature, in the context of the Frequentist versus Bayesian debate [35].

Another flavor of effect size found in the literature is the Cohen’s d statistic, which was intended to allow one to compare and combine experiments in a similar field but using different outcomes measurements. The essential idea is to standardize the magnitude of the effect through normalizing by some measure of variance. However, within a single experiment (where it never makes sense to use Cohen’s d), and in clinical reasoning, the phrase effect size refers to the absolute magnitude of the observed effect. Importantly, when the variances are proportional, it can be that Cohen’s d classifications are insensitive to the absolute versus relative effect problem. For example, a 10 % increase in mortality between groups with a pooled variance of 5 % has the same “effect size” as a 0.1 % increase with a pooled variance of 0.05 %. The qualitative gradation of effect size (small, medium, large) was devised by Cohen to apply to the social science context, not to biomedical research, where effect sizes can be larger, variances can be smaller, and there is a different philosophy behind the concept of effect size.

Another topic of importance to analysis of sleep physiology relates to reporting of sleep-wake stage architecture. Measurements of time spent in sleep and wake states are expected to anti-correlate simply because they are mutually exclusive, which can present interesting interpretation challenges. Similarly, REM and NREM are mutually exclusive and thus anti-correlated components of total sleep time. We recently explored the potential for embedded correlations, such that differentiating spurious versus meaningful correlations, for example related to a sleep stage, which itself

correlates with total sleep time, is not straightforward [20]. Experimental sleep deprivation necessarily involves increasing time awake, which may be associated with different physiology, independent of the increased stress that often accompanies the intervention. Stage-specific deprivation protocols are often associated with “collateral damage,” whether accomplished with pharmacology or with manual interruptions of sleep. For example, consider using an antidepressant to “suppress REM,” which has been described in the literature—one would not want to trivialize the litany of other antidepressant effects on neurochemistry. Similarly, consider an acoustic protocol designed to provide arousing stimulation in real time whenever N3 sleep is observed to decrease its occurrence. This could easily be shown to decrease the time spent in N3, but the intervention also may increase N1 and N2, increase the arousal index, or cause episodic cardiac changes, among perhaps more subtle neurophysiologic changes. Even referring to such a protocol as “N3-deprivation” implies these numerous other correlated changes are not relevant. Being mindful of these inferential topics is critical whether perusing the growing literature or contributing to it.

Conclusion

Reviewing any field of medicine is as much about celebrating progress as scrutinizing potential points of vulnerability. The intention of this book is to capture the breadth and depth of research into the health and performance consequences of sleep deprivation, in hopes of reducing the portion of the literature at risk for being later exposed as folly.

References

1. Dement WC, Vaughan CC. The promise of sleep: a pioneer in sleep medicine explores the vital connection between health, happiness, and a good night's sleep. New York: Delacorte; 1999.
2. Kroker K. The sleep of others and the transformations of sleep research. Toronto: University of Toronto Press; 2007.
3. Ekirch AR. At day's close: night in times past. 1st ed. New York: Norton; 2005.
4. Horne J. The end of sleep: 'sleep debt' versus biological adaptation of human sleep to waking needs. *Biol Psychol.* 2011;87(1):1–14.
5. Anderson C, Horne JA. Do we really want more sleep? A population-based study evaluating the strength of desire for more sleep. *Sleep Med.* 2008; 9(2):184–7.
6. Ioannidis JP. Why most published research findings are false. *PLoS Med.* 2005;2(8):e124.
7. Rattenborg NC, Lesku JA, Martinez-Gonzalez D, Lima SL. The non-trivial functions of sleep. *Sleep Med Rev.* 2007;11(5):405–9; author reply 411–7.
8. Rial RV, Nicolau MC, Gamundi A, et al. The trivial function of sleep. *Sleep Med Rev.* 2007; 11(4):311–25.
9. Grandner MA, Patel NP, Gehrman PR, Perlis ML, Pack AI. Problems associated with short sleep: bridging the gap between laboratory and epidemiological studies. *Sleep Med Rev.* 2010;14(4):239–47.
10. Buysse DJ, Grunstein R, Horne J, Lavie P. Can an improvement in sleep positively impact on health? *Sleep Med Rev.* 2010;14(6):405–10.
11. Epstein LJ, Kristo D, Strollo Jr PJ, et al. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *J Clin Sleep Med.* 2009;5(3):263–76.
12. Benbadis SR, Mascha E, Perry MC, Wolgamuth BR, Smolley LA, Dinner DS. Association between the Epworth sleepiness scale and the multiple sleep latency test in a clinical population. *Ann Intern Med.* 1999;130(4 Pt 1):289–92.
13. Chervin RD, Aldrich MS. The Epworth Sleepiness Scale may not reflect objective measures of sleepiness or sleep apnea. *Neurology.* 1999;52(1):125–31.
14. Gottlieb DJ, Whitney CW, Bonekat WH, et al. Relation of sleepiness to respiratory disturbance index: the Sleep Heart Health Study. *Am J Respir Crit Care Med.* 1999;159(2):502–7.
15. Eiseman NA, Westover MB, Mietus JE, Thomas RJ, Bianchi MT. Classification algorithms for predicting sleepiness and sleep apnea severity. *J Sleep Res.* 2012;21(1):101–12.
16. Budhiraja R, Budhiraja P, Quan SF. Sleep-disordered breathing and cardiovascular disorders. *Respir Care.* 2010;55(10):1322–32. discussion 1330–1322.
17. Brown LK. Quantum physics and polysomnography: can we prevent the act of measuring sleep from changing sleep? *J Clin Sleep Med.* 2005;1(2):133–5.
18. McCall C, McCall WV. Objective vs. subjective measurements of sleep in depressed insomniacs: first night effect or reverse first night effect? *J Clin Sleep Med.* 2012;8(1):59–65.
19. Horne J. Sleep debt—Where is the answer—in or outside the laboratory? *Biol Psychol.* 2011;87(2):314–5; author reply 316.
20. Bianchi MT, Wang W, Klerman EB. Sleep misperception in healthy adults: implications for insomnia diagnosis. *J Clin Sleep Med.* 2012;8(5):547–54.

21. Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose–response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*. 2003;26(2):117–26.
22. Vgontzas AN, Fernandez-Mendoza J, Liao D, Bixler EO. Insomnia with objective short sleep duration: the most biologically severe phenotype of the disorder. *Sleep Med Rev*. 2013;17(4):241–54.
23. Bliwise DL, Young TB. The parable of parabola: what the U-shaped curve can and cannot tell us about sleep. *Sleep*. 2007;30(12):1614–5.
24. Cappuccio FP, Cooper D, D’Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *Eur Heart J*. 2011;32(12):1484–92.
25. Bonnet MH, Arand DL. Hyperarousal and insomnia: state of the science. *Sleep Med Rev*. 2010;14(1):9–15.
26. Shekleton JA, Rogers NL, Rajaratnam SM. Searching for the daytime impairments of primary insomnia. *Sleep Med Rev*. 2010;14(1):47–60.
27. Daan S, Spoelstra K, Albrecht U, et al. Lab mice in the field: unorthodox daily activity and effects of a dysfunctional circadian clock allele. *J Biol Rhythms*. 2011;26(2):118–29.
28. Harvey AG, Tang NK. (Mis)perception of sleep in insomnia: a puzzle and a resolution. *Psychol Bull*. 2012;138(1):77–101.
29. Fichten CS, Creti L, Amsel R, Bailes S, Libman E. Time estimation in good and poor sleepers. *J Behav Med*. 2005;28(6):537–53.
30. Bliwise DL, Friedman L, Yesavage JA. Depression as a confounding variable in the estimation of habitual sleep time. *J Clin Psychol*. 1993;49(4):471–7.
31. Chaput JP, Despres JP, Bouchard C, Tremblay A. Longer sleep duration associates with lower adiposity gain in adult short sleepers. *Int J Obes (Lond)*. 2012;36(5):752–6.
32. Huck S. *Statistical misconceptions*. New York: Routledge; 2008.
33. Huff D. *How to lie with statistics*. New York: W. W. Norton; 1993.
34. Cohen J. The earth is round ($p < 0.05$). *Am Psychol*. 1994;49(12):997–1003.
35. Westover MB, Westover KD, Bianchi MT. Significance testing as perverse probabilistic reasoning. *BMC Med*. 2011;9:20.