Hyperarousal as the Basis for Insomnia: Effect Size and Significance

Comment on Varkevisser M; Van Dongen HPA; Kerkhof GA. Physiologic indexes in chronic insomnia during a constant routine: evidence for general hyperarousal? SLEEP 2005;28(12): 1588-1596.

Michael H. Bonnet, PhD

Dayton Department of Veterans Affairs Medical Center and Wright State University School of Medicine, Dayton, OH

IN RECENT YEARS NUMEROUS STUDIES HAVE SHOWN THAT PATIENTS WITH PRIMARY INSOMNIA HAVE SIGNIFICANTLY ELEVATED LEVELS OF AROUSAL compared to controls in several physiological domains including cardiac variables;1-3 whole body and brain metabolic variables;4,5 EEG spectral measures;6-10 immune function;11 and hormones including cortisol,12-15 melatonin,16,17 norepinephrine11 and ACTH.15 Occasional studies have reported only non-significant or trend changes in these measures (cortisol;16 beta EEG18). However, the number and breadth of significant findings assure us that these patients are physiologically distinct from matched controls.

Given these numerous positive findings, how does one explain the apparently negative results reported by Varkevisser et al in this issue of Sleep in an examination of cardiac measures and cortisol in a constant routine design in insomnia patients and controls? Of course, we all know that the null hypothesis can never be proven. However, beyond that, an occasional examination of the relationship between what we see, what our significance test tells us, and the level of the underlying effect size may help us both understand what at first appears to be controversy in the literature and better evaluate research findings.

In an experiment in which the active hypothesis is not upheld, one first examines the methodology of the study to determine if significant omissions, confounds, or other problems could have obscured the results. Perhaps the most common problem in expensive and labor intensive sleep research designs is collecting data from an insufficient number of subjects. The advent of power analysis has made determination of sample size less of an issue, but probability can always catch up with us. The cardiac data presented by Varkevisser et al provide a wonderful example. The statistical tests for all three cardiac variables presented by Varkevisser et al in their figures 2a, 2b, and 2c were not statistically significant. However, even a cursory examination of these figures (what we usually call the “eye trauma” test) reveals that heart rate was consistently elevated by 4-6 beats per minute in 9 of 9 observation periods throughout the 24-hour experiment in insomnia patients compared with controls. Such a finding is similar to flipping heads 9 consecutive times with a fair coin and has the exact binomial probability of p = .002.19 Data from the other two cardiac variables are equally impressive with binomial probabilities of .0196 and .002 respectively. The authors also calculated effect size for two of the three cardiac variables and these were reported as 1.04 (for heart rate) and 0.62.

Effect size is commonly graded as small (values around 0.2), medium (values around 0.5) and large (values around 0.8).20 For example, Cohen20 cites a study of the difference in IQ level between college graduates and students with a 50-50 chance of passing in high school as an illustration of an effect size of 0.8. This means that the effect size for the “non-significant” 4-beat increase in heart rate found by Varkevisser et al is 25% larger than the previously mentioned difference in IQ. However, finding significance is also based on chance. With an effect size of 1.0 and a sample size of 11 insomnia patients, the odds of Varkevisser et al obtaining a statistically significant result at the .05 2-tail level is 61% (i.e., despite a real and large effect, a statistically significant test will not be found 39% of the time). To increase the odds of statistically identifying such a real effect requires a larger sample size. In the current example, Cohen,20 for example, indicates that a sample size of 22 insomnia patients would have increased the odds of finding a significant result for heart rate to 90%.

Another means of identifying real differences is study replication. Two previous studies have documented heart rate in insomnia patients and matched controls during periods of wakefulness.1,2 Both found statistically significant results, with insomnia patients having increased heart rates of 6 and 7 beats per minute respectively. If the data from the current study and the two prior studies are combined in a meta-analysis, the resulting t-value, based upon 47 insomnia patients and 50 controls is t96 = 3.476 (p < .001) with combined means of 69.8 and 64.1 bpm. Looking at the individual study data, the mean heart rate for waking insomnia patients ranged from 68.8 – 70.5 bpm (the Varkevisser et al mean was 69.5) and the heart rate for waking controls ranged from 61.7 – 65.4 (the Varkevisser et al mean was 65.4). This suggests that the slightly smaller difference between groups in the current study compared to previous studies probably came from the control group.

Small N studies with between group comparisons are crucially dependent upon careful subject selection to maximize group differences and minimize variability. For example, in one heart rate study,2 subjects were literally paired based upon sex, age, and weight to control a few of the parameters that can affect heart rate. All three of the studies cited above required that insomnia patients demonstrate an EEG sleep efficiency of less than 85% on screening nights to qualify to participate. However, only the two studies showing a significant heart rate difference between the groups also required that normal controls demonstrate their normal sleep by having a sleep efficiency greater than 90% on a screening night.1,2 Sleep efficiency in those two studies was 91% and 93% in the control group on the experimental night.
Sleep norms for 40-year old individuals indicate a typical sleep efficiency of about 93% with a standard deviation of about 4%. Unfortunately, in the study by Varkevisser et al in this issue, the normal control group was not required to demonstrate a 90% sleep efficiency prior to admission to the study, and their actual sleep efficiency was 82% (i.e., the average control sleeper would have qualified to have been in the insomnia group). Inclusion of even one or two poor sleepers in the control group, obviously, would have increased the baseline heart rate (and variability) in that group, as indicated by the group mean, and decreased the odds of finding a statistically significant difference.

Another difference between the current study and some of the earlier studies is that the current study was done in a constant routine environment that did not allow sleep. This might have been an environment where there was more stress than typically seen in a sleep environment. Two previous studies have looked at stress during cognitive performance in insomnia patients compared to controls, and both have reported significantly increased heart rate in the insomnia patients. However, in one study, the control subjects had an increase in heart rate on the second study night and also had increased sleep latency (not found in the insomnia patients compared to their previous nights). Such data imply that the stress of the constant routine situation might have had more impact on the control subjects than on the insomnia subjects and could also have accounted for their higher than expected mean heart rate.

Statistical significance is a probability event, and that implies that there will be occasional losers. Inordinate concern with protecting from a spurious positive finding can result in the baby being thrown out with the bathwater. Sample size, replication, subject selection, and attention to effect size can almost always improve our focus.

REFERENCES