

Panel Discussion: Current Status of Measuring Sleepiness

Moderator: Janet M. Mullington, Ph.D.

Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA

Panelists: Charles A. Czeisler, M.D., Ph.D.¹; Namni Goel, Ph.D.²; James M. Krueger, Ph.D.³; Thomas J. Balkin, Ph.D.⁴; Murray Johns, Ph.D.⁵; Paul J. Shaw, Ph.D.⁶

¹*Division of Sleep Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA;* ²*Division of Sleep and Chronobiology, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA;* ³*WWAMI Medical Education Program, Sleep and Performance Research Center, Washington State University, Spokane, WA;* ⁴*Department of Behavioral Biology, Walter Reed Army Institute of Research, Silver Spring, MD;* ⁵*Optalert Pty Ltd., Epworth Sleep Centre, Melbourne, Victoria, Australia;* ⁶*Department of Anatomy and Neurobiology, Washington University, St. Louis, MO*

Dr. Mullington (Moderator): I think it has been a very interesting morning session and today the panel discussion is entitled, "Current Status of Measuring Sleepiness." To start, perhaps we can go through our speakers and hear their assessment of where we stand on our understanding of potentially two kinds of biomarkers, one being a roadside, safety type biomarker and another that is an indicator of some aspect of a person's health. Perhaps if we could start with Dr. Czeisler.

Dr. Czeisler: I'm particularly delighted that we have Dr. Murray Johns here from Australia, who as many of you know developed the Epworth Sleepiness Scale, and who has been working on the evaluation of a technology which he calls the Optalert method for looking at instantaneously drowsiness. This is very much related to the whole question of how do we identify a biomarker for sleepiness. Because it's interesting, as Dr. Balkin was alluding to earlier, that sleepiness is a very transient and evanescent phenomenon. When people are either chronically sleep deprived or totally sleep deprived, they describe it coming on in waves where they're overwhelmed by sleepiness. Then, without getting any further sleep, suddenly they get a little bit better; they seem less affected by it. There has been speculation that it's related to the basic rest-activity cycle, as Nathaniel Klideman called it, but you can have these episodes of profound drowsiness, where you can't keep your eyes open, and of course this is when you're most vulnerable to the drowsy-driving crashes. Dr. Johns has been trying to develop a technology that can instantly evaluate this. I would be interested to hear what he has to say about it.

Dr. Mullington: I think we will hear from Dr. Johns at the end of the brief assessments by the speakers and then we can open it up for discussion. So, Dr. Goel.

Dr. Goel: I think that, in terms of a health biomarker, that was one of the questions, it seems like we're not anywhere close to getting there from what the speakers earlier today talked about. In terms of a behavioral or perhaps a genetic biomarker, I also think that a lot of work still needs to be done although we have some hints of tests that could track sleepiness. I'll have more to say later.

Dr. Krueger: For health biomarkers, I think we'll probably talk a lot more about that tomorrow when we talk about chemis-

try. In follow up with what some of speakers said this morning, I think we can be optimistic. The difficulty is time and money to do the studies because they are very expensive. In terms of the behavioral biomarker and an instantaneous biomarker, Dr. Balkin alluded to near infrared imaging; one can do real-time NIR imaging of your pre-frontal cortex, and know the status of your pre-frontal cortex in real time. There are some people in Washington, where I work that are doing this kind of work. I joke with the students, saying that they're going to come in when I give an exam and they're going to be hooked up with near infrared detectors. Then, they're going to say, "Sorry, Dr. Krueger, I can't take your exam. My brain's not working at peak performance today." That's just one of many possibilities. I think there's a reason to be optimistic but, again, it takes time and money and a lot of expertise to do these things.

Dr. Balkin: I'm generally optimistic about things that I don't really understand and then the more I understand, the less optimistic I become. I'm generally optimistic about the possibility of a health sleepiness biomarker; something that will get at whether people are getting enough sleep in order to maintain proper metabolic panels and so on. I guess I am optimistic about a sleep behavioral performance biomarker as well. I think that there are ways to test for sleepiness on a moment-to-moment basis and they're fairly reliable and they're interpretable. But, in keeping with my tendency to not talk about things that they want me to talk about but what I want to talk about, I'd like to also add another comment on subjective sleepiness. In the laboratory, just anecdotally, a question was asked regarding how people rate themselves. In our data, what I showed you suggested that subjects rated themselves based on how they felt, pretty much, the day before. Our subjects were sleep restricted. However, even though their performance had not returned to normal, they rated themselves as normal subjectively because they felt so much better than they did the day before. They didn't rate themselves against any particular scale. However, I found that they do rate themselves against each other. They all, no matter what type they are, will say the same thing. We ran them in groups of four, and when we asked individually, "How are you doing?" they would say,

“Well, you know, I’m doing pretty good but these other three guys, they’re in real trouble.”

Dr. Johns: Can I just say at the outset how pleased I am to have this opportunity to come here, all the way from Melbourne, to talk about sleepiness and drowsiness. I’m just beginning to recover from my sleep deprivation and phase change. In discussions like this, it is always helpful to have some idea of what we *think* you’re talking about. We have heard the word sleepiness used in about six different ways today without explanation and perhaps even without recognition that they are not all the same. I don’t agree with several of them, so, we have a problem here. We need to decide what it is we’re talking about. If you go to an English language dictionary, the adjective sleepy is synonymous with drowsy. So, in a long-standing and traditional sense, the state of sleepiness is synonymous with the state of drowsiness. However, about 20 or 30 years ago, we in sleep medicine began to use the word sleepiness in another sense, meaning sleep propensity. Since then, various ways of defining sleepiness have evolved, none of them very well discussed or used consistently. About 20 years ago I developed the Epworth Sleepiness Scale (ESS). That is a subjective measure of sleep propensity in a variety of different situations. It is not a measure of drowsiness, in the sense that the Karolinska Sleepiness Scale is. The multiple sleep latency test (MSLT) also measures sleep propensity, but in only one particular test situation. The maintenance of wakefulness test (MWT) measures sleep propensity in a different situation. Each of these tests which purports to measure sleepiness as sleep propensity is actually measuring something different. We simply do not have a unitary concept or a gold standard measure of sleep propensity as a general characteristic of someone in their daily life. Hopefully, we can more easily distinguish the two basically different meanings of the word sleepiness, as currently used – drowsiness and sleep propensity.

My recent focus of attention has been on the measurement of different levels of drowsiness, the intermediate state between alert wakefulness and sleep. There are many physiological markers of drowsiness that haven’t been mentioned so far today. I have a list of 25. They are all derived from ocular dynamics, the way our eyes and eyelids move when we’re doing things, intending to remain alert but sometimes becoming drowsy. This interest of mine developed through a desire to understand and perhaps to solve the problem of drowsy driving. What I was after was a measure of drivers’ drowsiness from minute to minute while they were driving. With that information, we could warn them when they’ve first showed the signs of drowsiness with increased risk of performance failure, i.e. driving off the road. We’ve come a long way in that. We use infrared reflectance oculography as a way of collecting information about drowsiness. That is like the EOG, but doesn’t have its disadvantages. You don’t have to attach electrodes, but wear a pair of special glasses instead. We now have a scale of drowsiness (the Johns Scale of Drowsiness, or JDS), based on a weighted combination of several ocular variables measured each minute. This has been calibrated against the relative risk of performance failure at a variety of different tasks, including driving after being sleep deprived.

Fitting in with what Dr. Balkin said earlier today, the best measures we have of drowsiness at a particular time are not the

mean values of our markers but their standard deviations. We have based our scale of drowsiness mainly on standard deviations, particularly the relative velocities of eyelid closing and reopening movements during blinks. Those measurements provide bio-markers of drowsiness related to the neuro-muscular functions of the eyelids controlled by reflexes. Those functions are highly controlled by the brain when we are alert, but become inhibited and more variable when we are drowsy. Of the 25 ocular biomarkers of drowsiness that we can measure, many are significantly intercorrelated. Many also show differences between subjects, with an effect-size that is comparable to that observed within subjects after sleep deprivation for 24-30 hours. These between-subject differences are much less for four of our 25 variables. Those four are standard deviations rather than means, calculated per minute. They form the basis of our algorithm for calculating JDS scores from minute to minute. The scores are quite variable, depending on what the person is doing (e.g. on the nature of the driving task at the time).

Dr. Mullington: I think we can open up to the audience. You can comment on anything that you’ve heard through the morning sessions in terms where we are in measuring sleepiness, or comment on definitions and operationalizing the problem. So if anybody would like to begin from the floor. Dr. Shaw.

Dr. Shaw: It seems like one of the first goals ought to be to find a chemical or behavioral task that’s reliably changed after a fixed period of sleep loss and then to begin validating what that construct actually means. Unless we have something we can measure, it’s all meaningless. We need defined variables, whatever they may be, and then we can begin asking more precise questions. So if I can reliably tell you that if I see something, call it “A,” and it reflects that somebody’s been awake for 24 hours or more, it becomes useful no matter what that A ultimately represents.

It’s also, I think, worth remembering that we might need a panel of tests. We talk about roadside uses of a biomarker. You are pulled over by the police and they think you’re drunk. They give you behavioral tests first. Touch your nose and then they give you a breathalyzer. More than one assessment might be required for this as well, where you need multiple tests in combination to come to a definitive conclusion.

Dr. Mullington: Would you agree that we need to define what is the behavior in terms of risk of behavioral failure? If what we’re talking about is a roadside test, this should be what a panel should predict.

Dr. Shaw: I want to know whether I can tell you reliably that you’ve been awake for 24 hours or more. Afterwards, we can decide whether or not that has ramifications in terms of failures. What is your risk of driving off the road if you’ve been awake for 24 hours? It doesn’t have to be 100% for us to not want people on the roads that have been awake for 24 hours.

Dr. Johns: We can do this today. We are doing it as we speak.

Dr. Mullington: Do you want to comment on the reliability of what you are doing, and give us a little bit more information about it?

Dr. Johns: I have to be careful I don’t promote my own commercial interests. We have several hundred drivers in South America, Australia and Canada who are driving trucks as we speak, and whose levels of drowsiness are being measured every minute. That information is provided directly to the drivers

and also transmitted in real-time to their managers and, via the web, back to us in Melbourne. This is providing a new way to manage the safety and efficiency of vehicles used in long-haul transport and in mines. JDS scores have been shown to be highly reliable in a test-retest sense and to be valid in the sense of being able to predict performance failure in a series of performance tests. We can predict driving off the road events from the pattern of ocular variables measured each minute. So, that's already being done.

Dr. Mullington: So then you have a time? You can link that to a duration of wakefulness? Those data? Those risks?

Dr. Johns: Yes, we do. A paper is currently being prepared on it. The drowsiness scores are closely related, in a statistical sense, to the duration of prior wakefulness and to circadian phase.

Dr. Czeisler: Thinking the way Senator Moore (Richard T. Moore, Massachusetts State Senator, Sponsor of Drowsy Driving Legislation) would think at this moment, the only problem is that most of the people are not wearing this. Ideally, you would have something where someone who is not had the forethought to be trying to monitor their state, but who has been irresponsible and is driving and is now pulled over. Then something could be measured at that point in an individual. That would be the ideal thing.

Dr. Johns: We're working on that too.

Audience Member: I have a short question and another short question. The first one goes to Dr. Johns. Can I wear your gadget and you can tell me how long I have been awake right now?

Dr. Johns: Yes, but you will have to recognize that hours of wakefulness is not a particularly good predictor of performance. Some people will have been awake for 30 hours and hardly show any impairment at all, as is the case with an elevated blood alcohol level. We're not really measuring hours of wakefulness as a variable. What we're measuring is the risk of performance failure associated with particular levels of drowsiness at the time.

Audience Member: Can I wear your gadget and can you tell me what is the probability that I can't walk back to my seat right now without calibrating the device?

Dr. Balkin: I'd just like to point out that, again, as Dr. Johns just said, the issue isn't how long you've been awake. We can put actigraphs on you and tell you how long you've been awake.

Audience Member: I'm interested in that you can do it without calibrating it. That's what I've become very, very interested in because you said there is a 20% biological variation or more. Still you say that, without calibrating, you can know where I sit on that 20%?

Dr. Johns: I didn't say there's 20% variation.

Audience Member: I saw the error bars there and, normally in biology, it's 20%.

Dr. Johns: Well, I don't accept that. No, not in this model.

Audience Member: But basically you can say that, without calibration, you can tell my status.

Dr. Johns: Yes.

Audience Member: I would like to sign up right away. My second question, if I may?

Dr. Mullington: I think we should take some other questions.

Audience Member: I am currently with the University of Maryland and, looking at the guest list, I suspect I may be the only lawyer in the room. As I understand it, this conference

originates from concerns related to drowsy driving, that is to say, how are we going to go about using this as evidence. The gold standard, of course, as Dr. Czeisler said, would be a breathalyzer for sleep and there's a reason why that is. This is a comment to the room of researchers as we move forward on this path, and figure out what biomarkers or how we can identify biomarkers. The reason why lawyers and policy makers like a breathalyzer for sleep is because when a car crash happens, police investigate. They try to figure out what happened. Their key job is to assemble, at least in the United States, a package of evidence to present to the prosecutor, who then makes a decision whether or not to prosecute under the existing laws. That decision is based on what sort of evidence they could present and what they can admit in court. As it stands now, without a breathalyzer, if you have all these different measures and different ways of trying to assess sleepiness or sleep propensity or whatever, there needs to be expert witnesses to testify concerning the reliability of these things. Lawyers can't make that call whether this is reliable evidence or not. They need to bring an expert. Even if you present a report that says, "Yes, these measures were identified from this defendant." You still need to bring in an expert to attest to that. So that's one of the challenges. Furthermore, the more consensus there is within the community, the better it looks in court. That's something to keep in mind as we move forward. Any comments that you might have to what I just said would be appreciated.

Dr. Balkin: I've got a comment. Actually, I just testified in a trial in Virginia. The question actually came down to whether he was sleepy. He admitted that he hadn't slept. This guy was driving a truck, fell asleep, crossed the centerline, and killed three people. The question came down to whether, given his sleep history and what he admitted, whether he should have known that this was likely. That is, whether he subjectively should have known that he was so sleepy that it constituted a danger to other people. Now, he was convicted but I'm not sure we actually showed that.

Audience Member: I think that, based on what many of the panelists have said, and Dr. Johns has clearly pointed this out, there are different ways of defining sleepiness, different contexts. What we've heard about today are things ranging from distinguishing people who are habitually sleepier from other people and people who respond better or worse to sleep loss. Additionally, there is context to their performance measures. One of the original challenges that Dr. Czeisler had for us was what I would call the forensic definition of sleepiness. I'm wondering, in that context, if what we need is not really a biomarker but just a black box. So, if what we're worried about is car crashes, why not have black boxes in cars that measure variance in performance, in other words, lane deviations and so forth that would give a reliable indicator of the person's actual performance. These would not depend on self-report and would not depend on a roadside test administered after the fact that has no chance of guessing at what the person's level of sleepiness was before hand.

Dr. Mullington: And maybe it is more than just whether or not they are sleepy? So if they are poor drivers?

Audience Member: Right. As several people have pointed out, who cares how long someone's been awake? I don't care about that. What I care about is if they're going to drive into me.

And I also don't care if it's because they're sleepy or because they're intoxicated or because they're just bad drivers. What I want to know is, what is the risk that they have? So, in that forensic context, again, is instrumenting the vehicle better than instrumenting the individual?

Dr. Balkin: Does anyone on the panel want to comment?

Panel Member: Well, yes. I think there's a lot of validity to that and, as you were talking, Tom, I was thinking about the Perclos, which measures percentage closure of the eye lid. It was actually validated, to the extent that it was validated initially, against lane deviations. It's probably easier to measure lane deviations than it is eye closure. Once they developed the lane-deviation technology, I wondered why did the Perclos just go away. But that's, of course, only for driving.

Audience Member: I'd like to address Professor Krueger, if I may. I liked your talk very much and you were the only one, as far as I remember, that spoke about the physics of sleepiness. Being a physicist, of course, I woke up at that point. You spoke about signal conduction, basically, electrical conduction. My question is very simple. In your opinion, what is the biophysics background to sleepiness?

Dr. Krueger: Simple questions, right? I think it's a simple question without an answer.

Audience Member: It's interesting because everybody has spoken either from a behavioral point of view or from a biochemical point of view. I believe that signal conduction may be very interesting to look at, perhaps. I don't think it goes to the state of the art, which perhaps this panel is more about today.

Dr. Krueger: I'll talk more tomorrow about our ideas of how bits and pieces of the brain can be asleep and awake simultaneously and show data related to that. I don't know if that's getting at what you want but it is electrophysiologically based and biochemically based. I'll talk about that tomorrow.

Dr. Mullington: I think we have time for one more question.

Audience Member: Much of discussion today has been about acute effects of sleepiness such as the effects on driving and cognitive function. I'd like the panel to comment a little bit about the more prevalent problem which is chronic sleep deprivation. Is that the same thing as sleepiness and the long-term health effects? It seems to me like we might be talking about two very, very different things.

Dr. Balkin: We really only recently started to look at chronic sleep restriction and there is some evidence that, to some extent, they may be different animals from acute sleep deprivation. That is to say, chronic sleep restriction may produce some effects that are different from acute sleep loss. There's good

physiological reason to think why that may be the case, in terms of adenosine receptor regulation and down regulation and with respect to the availability or release of extra cellular adenosine. That work, of course, has been done by Drs. Strecker and McCarley and others at Harvard. But the short answer to your question is, yes, there are probably differences between the two. If you recall, in the figure that Dr. Goel showed of the study done at Walter Reid, there were five consecutive days of sleep restriction and I think 11 or 12 days at the University of Pennsylvania. If you recall, although she didn't make anything of it, there were two areas that were shaded. One was shaded lightly and one was shaded darkly. The lines went into each of those areas at different points representing different days of sleep restriction. What those lines actually represented was the amount of sleep restriction that was equivalent to 24 and 48 hours of total sleep deprivation and performance on the PVT. In terms of PVT performance, yes, you could definitely get to the similar sorts of performance deficits with chronic sleep restriction. It just takes longer.

Dr. Czeisler: I would also point out that there actually can be an interaction between them. Drs. Daniel Cohen and Elizabeth Klerman at Harvard this past January, published a study showing that if you take subjects and put them on a regime where they're only getting five hours of sleep every 24 hours within a week of being on that restricted schedule the impact of being awake for more than 24 hours increased ten-fold. When they were in an adverse circadian phase during the biological night, the impairment was ten times worse than staying awake for the same number of hours when they were well rested. So it's not one plus one equals two. It's one plus one, in this case, equals ten at the adverse circadian phase. That was even after a ten-hour episode of recovery sleep. So, the ten-hour episode of recovery sleep was insufficient to overcome the vulnerability to acute sleep loss that had built up from the chronic sleep restriction.

Dr. Balkin: As many people in this room know, this is the problem we are, in the military particularly, interested in. Because the military population whose performance we are trying to sustain, is characterized by chronic sleep restriction punctuated by periods of total sleep loss.

ACKNOWLEDGMENT

Edited by Stuart F. Quan, M.D., Conference Chairperson and Supplement Editor, Division of Sleep Medicine, Harvard Medical School, Boston, MA. Editing of the conference proceedings was supported by HL104874.