Melatonin and Human Chronobiology

A.J. Lewy

Department of Psychiatry, Oregon Health & Science University, Portland, Oregon 97239-3098

With the development of accurate and sensitive assays for measuring melatonin in plasma and saliva, it has been possible to advance our understanding of human chronobiology. In particular, the dim light melatonin onset (DLMO) is expected to have an increasingly important role in the diagnosis of circadian phase disorders and their treatment with appropriately timed bright light exposure and/or low-dose melatonin administration. The phase angle difference (PAD) between DLMO and mid-sleep can be used as a marker for internal circadian alignment and may also be used to differentiate individuals who are phase advanced from those who are phase delayed (a long interval indicates the former and a short interval indicates the latter). To provide a corrective phase delay, light exposure should be scheduled in the evening and melatonin should be administered in the morning. To provide a corrective phase advance, light exposure should be scheduled in the morning and melatonin should be administered in the afternoon/evening. The study of patients with seasonal affective disorder (SAD), as well as individuals who are totally blind, has resulted in several findings of interest to basic scientists, as well as psychiatrists and sleep specialists.

INTRODUCTION: THE GCMS MELATONIN ASSAY

The longer duration of pineal melatonin production during winter nights compared to the summer is a time-of-year signal essential for the regulation of seasonal rhythms, which are ubiquitous, particularly in mammals, for example, seasonal reproductive cycles. Both long- and short-day breeders rely on melatonin to signal the “biological night.” This function was elucidated by measuring melatonin production in extracts from whole pineal glands. Human studies, however, had to wait until plasma levels of melatonin could be reliably measured: The gas chromatographic–negative chemical ionization mass spectrometric (GCMS) assay using a deuterated internal standard achieved the requisite accuracy, sensitivity, and precision for measuring melatonin in humans (Lewy and Markey 1978). This assay was instrumental in settling the controversial issue of reports of extrapineal sources contributing to melatonin levels in the circulation (Ozaki and Lynch 1976): Exclusive derivation of melatonin from the pineal validated the use of circulating melatonin levels (Lewy et al. 1980a; Neuwelt and Lewy 1983) as an indicator of pineal melatonin production, of the timing of the endogenous circadian pacemaker (located in the suprachiasmatic nucleus of the hypothalamus), and of the effects of the light/dark cycle (mediated by the retinohypothalamic tract). This assay also directly and indirectly enabled the eventual development of widely used RIAs (radioimmunoassay) with sufficient specificity for measuring melatonin in human plasma and saliva.

DISCUSSION

Suppression of Nighttime Melatonin Production by Light

The GCMS assay was also used to show that nighttime human melatonin production could be suppressed by light, provided it was sufficiently intense; light of intermediate intensity produced an intermediate amount of suppression (i.e., the response was not binary) (Lewy et al. 1980b). Ordinary intensity light is usually insufficient in humans (Arendt 1978; Wetterberg 1978), although all animals suppress with very little light (Illnerová et al. 1978; Illnerová 1979). We speculated that exposure to sunlight renders humans less sensitive to light (Lewy et al. 1980b). In fact, squirrels tested immediately after being caught in the wild require bright light for melatonin suppression (Reiter et al. 1981). Subsequent human studies have documented the relationship between history of prior exposure and sensitivity to light (Hebert et al. 2002; Smith et al. 2004). However, restricting ambient exposure to dim light for 1–2 days in the constant routine protocol can exaggerate melatonin suppression and circadian phase-shifting responses (Boivin et al. 1996; Zeitzer et al. 2000). Although all intensities should ideally be taken into account when relating circadian phase to the ambient light/dark cycle, intensities of 2,000–10,000 lux are standard for light treatment.

The Melatonin Suppression Test

The melatonin suppression test (MST) was first applied to demonstrate supersensitivity in patients who were actively manic or depressed (Lewy et al. 1981). Subsequently, 500 lux was shown to cause 50% suppression in healthy control subjects, whereas euthymic manic-depressive patients suppressed almost completely at this intensity (Lewy et al. 1985d), an intensity that is ideal for identifying supersensitivity. A version of the MST has been used in blind people with no conscious light perception: 10,000 lux causing 33% suppression has been proposed by some researchers as a way to inform ophthalmologists about the advisability of therapeutic enucleation (Czeisler et al. 1995), in that they interpret a positive test (i.e., suppression) to mean—even in cases of total and irreversible loss of vision and conscious light perception—that bilateral enucleation should not be done because the eyes should...
still be able to functionally mediate entrainment. However, several objections can be raised about this use of the MST. The arbitrary choices of the 10,000-lux light intensity and 33% suppression threshold do not take into account the fact that blind people are often exposed to sunlight as bright as 100,000 lux. Furthermore, because of the possibility of photoreceptors becoming up-regulated following blindness, people with no conscious light perception may still be capable of entraining to even low-intensity light. Another objection relates to safety: A blind person staring at a bright light fixture emitting 10,000 lux runs the risk of possibly damaging their few remaining photoreceptors, particularly if they have become up-regulated. Another objection is that some blind people with a positive MST nevertheless free-run (indeed, some sighted people free-run) (McArthur et al. 1996), whereas some blind people with a negative MST are naturally entrained to the 24-hour day (even some who are bilaterally enucleated; see below). A different type of MST that satisfies these objections must be configured before its use in blind people can be recommended.

Plasma Melatonin Profiles in Totally Blind People

The study of 24-hour melatonin rhythms in totally blind people and the exploration of what might be the consequences of light deprivation also resulted from the finding that light can suppress melatonin production in sighted humans. Previously, human chronobiology researchers had concluded that the light/dark cycle was relatively unimportant compared to social cues (Wever 1979). However, we described abnormally phased 24-hour plasma melatonin profiles in six of ten blind subjects studied in December, 1979 and January, 1980 (Lewy 1981). Fifteen months later, two of them were studied weekly on four occasions: The subject who was bilaterally enucleated appeared to be entrained but at a very abnormal phase, whereas the other subject was found to have a free-running melatonin rhythm with a circadian period (tau) of 24.7 hours (Lewy and Newsome 1983). Accordingly, we proposed that blind people can be categorized into three types: Those who are entrained at the normal phase, those who are entrained at an abnormal phase, and those who are free-running. Smith et al. (1981) reported that serum melatonin values in four blind people were greater at 1400 (86–142 pg/ml) than at 2300 (64–72 pg/ml). However, these data should be interpreted with care, because high daytime levels were routinely found with the nonspecific RIA's in use at the time (Smith et al. 1977). It should also be noted that Lynch et al. (1975) were the first to describe abnormal (urinary) melatonin levels in blind people (because of space constraints, historical notes will be limited in this volume).

Circadian Phase-shifting Effects of Light

Because sunlight is usually brighter than indoor light, two other implications of the melatonin suppression finding were that sighted people might have biological rhythms cued to the natural light/dark cycle (relatively unperturbed by ordinary-intensity room light) and that bright artificial light could be used to experimentally, and perhaps therapeutically, manipulate these rhythms. The first rhythms to be tested, however, were not thought to be circadian. Just after Kripke et al. (1983) began to treat major depressive disorder with morning light, based on his “critical interval” theory, we treated a patient with winter depression (Lewy et al. 1982), a disorder previously unknown to us and subsequently described as seasonal affective disorder, or SAD (Rosenthal et al. 1984). He responded after several days of receiving 2000 lux scheduled at 6–9 a.m. and at 4–7 p.m., so as to lengthen his perceived photoperiod. Subsequently, a group of patients were successfully treated under more controlled conditions (Rosenthal et al. 1984). SAD is discussed in detail later in this chapter.

Regarding circadian effects, Wever reported in 1983 that bright light scheduled throughout the photoperiod could increase the range of entrainment to a gradually lengthening T cycle (Wever et al. 1983). In the same year (Lewy et al. 1983), we proposed that bright light could be used according to our hypothesized human phase-response curve (PRC) to treat circadian phase disorders, such as delayed sleep phase syndrome (DSPS). At the time, DSPS was treated by scheduling sleep 3 hours later each day (termed “chronotherapy”; Czeisler et al. 1981), which was based on a two-pacemaker model (Kronauer et al. 1982). Their thinking was that there were two endogenous circadian pacemakers, one for the sleep/wake cycle (located in the suprachiasmatic nucleus) that could directly entrain a separate pacemaker for temperature (thought to be located elsewhere; Kronauer et al. 1982). This model explained internal desynchronization observed in temporal isolation, but so did the one-pacemaker model proposed by Eastman (1982). In any event, we began to treat patients with DSPS with morning bright light exposure (Lewy et al. 1983), an intervention that continues to be preferred over chronotherapy (Wright et al. 2006). Our hypothesized bright light PRC was also the basis for scheduling a sunlight exposure for two subjects who had flown across nine time zones (Daan and Lewy 1984). One subject avoided sunlight for the first 3 hours in the morning, obtaining it after 10 a.m., which according to our hypothesized light PRC would be the beginning of the advance zone before it began to adjust to the new time zone: The temperature rhythm quickly advanced nine hours. However, the subject who obtained sunlight exposure beginning at 7 a.m., thus stimulating the delay zone, shifted in that direction and did not adapt to the new time zone even after 2 weeks.

The Clock-Gate Model, the Light PRC, and the DLMO

In 1984 and 1985 (Lewy et al. 1984, 1985a), we published a test of our “clock-gate” model (Lewy 1983; Lewy et al. 1985b) for determining how light regulates the melatonin circadian rhythm in humans. Holding the sleep/wake cycle constant in four healthy control subjects also enabled the test of our proposed PRC to light (Lewy et al. 1983) and the first use of the DLMO. The DLMO is now
acknowledged as the best marker for circadian phase position in humans, with a standard deviation that is less than half of the sampling interval (Klerman et al. 2002a).

Some researchers believe that the DLMO marks the phase of only one oscillator for melatonin production (Wehr et al. 1993; Parry et al. 1997; Benloucif et al. 2005), based on the elegant two-oscillator model proposed by Illnerová and Vaněcek (1982). This model consists of an evening oscillator cued to dusk that controls the onset of melatonin production and a morning oscillator cued to dawn that controls the offset of melatonin production, which explains the finding in rats that a very short pulse of light in the middle of the night flattens melatonin levels for a few days. However, humans do not respond to light in this way (Vondrašova-Jelínková et al. 1999).

The clock-gate model is the most parsimonious explanation for changes in melatonin duration measured under naturalistic ambient light conditions. Only the two-oscillator model can explain changes resulting from a light/dark cycle preceding measurement under dim light conditions. However, these changes (best assessed using the melatonin synthesis offset, or SynOff, rather than the DLMOFF to mark the putative second oscillator) (Lewy et al. 1999) are small and can only be demonstrated using exotic light/dark cycles (Wehr et al. 1993). Therefore, if there are two oscillators for human melatonin production, they are tightly coupled under most circumstances.

The interventions required for minimizing masking effects when measuring phase markers can alter phase. This is one reason why we do not reduce light intensity below 10–30 lux, why we do not begin dim light before 5 p.m. (about 1 hour before the earliest expected DLMO), and why we obtain DLMOs no more frequently than once per week (Lewy et al. 2006a). Under these circumstances, if there is a same-day phase advance in the DLMO due to the dim light intervention, it is thought to be small. The phase-advancing effects of the constant routine have not been studied since an earlier report (Czeisler et al. 1985). The effects of several days of dim light, as well as other interventions that are part of the forced desynchrony (FD) protocol (Czeisler et al. 1999), could be even more profound in sighted individuals (see below).

The DLMO was used to compare the phase-shifting effects of morning versus evening light (Lewy et al. 1984, 1985a, 1987a). In our studies, light was scheduled no earlier than wake time and no later than bedtime, so as not to interfere with sleep. Subsequently, four complete PRCs testing exposure at other times of the day and night were published (Honma and Honma 1988; Czeisler et al. 1989; Wever 1989; Minors et al. 1991). The most detailed was also the most controversial (Czeisler et al. 1989), in that claims were made that two daily light pulses could suppress the amplitude of the endogenous circadian pacemaker. Suppression of circadian amplitude is at present not an area of much research or clinical interest, which is one reason why measurement of melatonin has replaced temperature as a research tool. The DLMO provides the most important information for assessing the circadian system.

Although the DLMO was originally measured in plasma, it is now more often measured in saliva (Voultsios et al. 1997). Saliva levels are routinely about one-third those of plasma, when they are low, such as for the DLMO. The most common operational definition of the DLMO is the interpolated time when levels continuously rise above 10 pg/ml in plasma (3 pg/ml in saliva). In low secretors, we use the lower thresholds of 2 pg/ml plasma (0.7 pg/ml saliva) (Voultsios et al. 1997). Saliva levels are routinely about 14 hours after wake time in entrained sighted people. PRCs are often plotted with the abscissa given in CT. When of the light PRCs (Czeisler et al. 1989) was published according to established conventions (Johnson 1990) (such as using the beginning of the light pulse as its phase reference point), the crossover times were at CT 6 and CT 18. In Figure 2, the crossover times are taken from this light PRC and the times for optimally and conveniently scheduling bright to cause phase advances and delays are taken from our work (Lewy et al. 1987a). Conversion of CT to clock times is convenient when comparing DLMOs that have been calculated using these two thresholds. Different thresholds are sometimes used for the same person, for example, if assay parameters vary; however, they are more often of use when comparing the DLMO of one person to another.

No matter what its clock time, the DLMO \(_{10}\) is designated CT 14. This is because the DLMO \(_{10}\) occurs on average about 14 hours after wake time in entrained sighted people. PRCs are often plotted with the abscissa given in CT. When one of the light PRCs (Czeisler et al. 1989) was published according to established conventions (Johnson 1990) (such as using the beginning of the light pulse as its phase reference point), the crossover times were at CT 6 and CT 18. In Figure 2, the crossover times are taken from this light PRC and the times for optimally and conveniently scheduling bright to cause phase advances and delays are taken from our work (Lewy et al. 1987a). Conversion of CT to clock times is also provided, based on an average wake time of 6 a.m. and an average DLMO \(_{10}\) of 8 p.m. Optimal times to administer melatonin based on our melatonin PRC in
sighted people (discussed below) (Lewy et al. 1992, 1998a) are also provided in this figure. Wake time (CT 0) is a convenient way to estimate the phase of these PRCs in sighted, entrained people, although the DLMO10 (CT 14) provides a more accurate estimate. Wake time is also not useful if a (sighted) individual has had a very recent change in their sleep/wake cycle, such as occurs in shift workers. People who have traveled across time zones, the pre-travel wake time can be used to estimate the phase of the PRCs on arrival at destination, which are expected to shift at least 1 hour per day until adaptation is complete (Lewy et al. 1995). With appropriately scheduled sunlight exposure and/or melatonin treatment, the rate of adjustment should be about 2–3 hours per day.

Fortunately, obtaining a DLMO does not interfere with normal sleep times, because it almost always occurs before habitual sleep onset, particularly the DLMO12. Given the above, it is not surprising that the DLMO has become the circadian phase marker of choice in research studies. The only masking influence that needs to be controlled is light. Posture, which once was thought to be a potential confound by some investigators (Deacon and Arendt 1994), does not seem to have much effect on the DLMO (Cajochen et al. 2003).

Given its usefulness, reliability, simplicity, and convenience, there is little doubt that the DLMO will become a standard medical test, particularly after more researchers and clinicians gain experience with home saliva collections. Kept cool, samples can be express-mailed back to the laboratory; after centrifuging the absorbent material, the saliva is immediately frozen. In blind people, the MO can occur during the hours of sleep; however, sampling every 1–2 hours between wake time and bedtime for a “wake time circadian phase assessment” will result in an MO and/or an MOFF, because the duration of active melatonin production is shorter than the duration of wakefulness. We can estimate the MO from the MOFF, either because on another occasion we have determined either the MO/MOFF interval for that individual or by assuming these intervals to be about 10 or 14 hours, respectively, using the MO12. Blind people have an easier time with collections, because the greatest challenge for sighted individuals is maintaining strict dim-light conditions. However, amber goggles that filter out blue light may prove to be helpful. Incidentally, the advantages of blue or blue-enhanced white light for causing phase shifts versus the risk of the “blue light hazard” will not be discussed further here. No doubt these topics, as well as that of circadian photoreception, will be considered in detail elsewhere in this volume.

### Entrainment of Blind People to Melatonin and the Melatonin Dose-response Curve

Before the landmark study of Redman, Armstrong, and Ng (1983), in which daily melatonin injections were shown to be able to entrain free-running rats, there was not much interest in the circadian effects of melatonin in mammals, despite the fact that in other animals (birds and reptiles), melatonin has profound circadian effects (Underwood 1986; Cassone 1990). In mammals, melatonin was thought to mediate primarily, if not exclusively, biological rhythms of the seasonal type (for review, see Arendt 1995). Humans, as well as some strains of rats, may be exceptional, in that melatonin may have more of a circadian effect than a seasonal one. In fact, it may be disadvantageous for a mammal to use melatonin for both circadian and seasonal timekeeping. Indeed, hamsters are more sensitive to the circadian phase-shifting effects of melatonin injections when they are perinatal and sexually incompetent (Davis and Mannion 1988).

The landmark study inspired Robert Sack and I to first investigate the effects of melatonin in BFRs (blind free-runners) (Sack et al. 1987). The same study may have also influenced the first two investigations of the phase-shifting effects of melatonin in sighted people (Arendt et al. 1985; Mallo et al. 1988; for review, see Lewy and Sack 1997). Although we published a case report showing entrainment of a blind person to melatonin in the early 1990s (Sack et al. 1990, 1991), it was not until the end of that decade that unequivocal entrainment (to 10 mg) could be demonstrated in a group of BFRs (Sack et al. 2000). In another study of the same number of subjects (Lockley et al. 2000), barely half entrained to 5 mg, leading the authors to conclude that starting the first bedtime dose as it happened to coincide with the advance zone of the melatonin PRC was critical for entrainment. However, extensive animal literature indicates that entrainment is as likely to (eventually) occur if the phase-resetting agent is started on the delay zone, which is what we have found in...
BFRs (Emens et al. 2003; Lewy et al. 2004a). When the dose was switched from 5 to 0.5 mg, the advance versus delay zone difference was less (Hack et al. 2003) and the authors continued to recommend against initiating treatment on the delay zone. Recently, however, they have indicated agreement with our thinking about this issue, at least with respect to low doses (Lockley et al. 2007).

Too high a dose may make entrainment less likely if there is too much spillover of melatonin levels onto the wrong zone of the melatonin PRC. When given on the advance zone, an advance will always occur, but its magnitude will be reduced due to spillover onto the delay zone. Therefore, some low doses are capable of entraining a BFR than some higher doses, particularly if the BFR has a long tau and needs a large daily phase advance in order to entrain (Lewy et al. 2002). Melatonin administration was originally given at bedtime (Lockley et al. 2000; Sack et al. 2000). However, this results in an entrained MO at a delayed phase (Lewy et al. 2001). Although entrained, they will complain of difficulty falling asleep and getting up in the morning, even though the high doses used (5–10 mg) would be expected to have soporific side effects. Hence, we now give low doses of melatonin to most BFRs around 6 p.m., which results in an entrained MO at 8 or 9 p.m. (or about 2–3 hours before desired sleep onset). Low doses (0.3–0.5 mg) will not cause unpleasant soporific effects between 6 p.m. and bedtime.

In any event, treatment of both blind and sighted people is based on the melatonin PRC. When melatonin is given between CT 6 and CT 18, it causes phase advances; when it is given between CT 18 and CT 6, it causes phase delays. There was some skepticism about the delay zone in our first report (Lewy et al. 1992), despite independent confirmation (Zaidan et al. 1994; Middleton et al. 1997), leading some researchers to suggest that melatonin could only cause phase advances (Wirz-Justice et al. 2002) and that melatonin’s main circadian effect in humans was to shorten tau (Arendt et al. 1997; Czeisler 1997). In a second study, we gave melatonin at more times during the night, leaving no doubt about the existence of a robust delay zone (Lewy et al. 1998a). Nevertheless, the tau-shortening effect of melatonin on human circadian rhythms remains an issue with some currency (Arendt 2006).

Once entrainment occurs, the time interval between the melatonin dose and the MO at steady state is the phase angle of entrainment (PAE), which correlates with tau: The longer the tau, the greater the PAE (Fig. 3) (Lewy et al. 2001). Given the relationship between tau and PAE, we have recommended that PAE be used to estimate tau (Lewy et al. 2003). However, in BFRs, measurement of tau in the field is relatively easy to do and is preferred. In sighted people, some investigators use the DLMO/sleep onset interval (melatonin/sleep interval, or MSI) to estimate the PAE in sighted people (Wright et al. 2005). We prefer to use the sleep offset/DLMO interval, which we have termed the ZT (zeitgeber time) of the DLMO. However, although DLMO ZT adheres to the traditional way of calculating PAE by using “lights on” for marking ZT 0, humans are exposed to a complex light-intensity contour that varies throughout the day. We will continue to use wake time as the phase reference for the light/dark zeitgeber, until a better way of calculating PAE in sighted people is found.

As mentioned above, for BFRs with taus greater than 24 hours, taking melatonin at about 6 p.m. results in an entrained MO at about 8–9 p.m. If a BFR with a tau less than 24 hours takes melatonin at 6 p.m., they will likely have an entrained MO at about 8–9 a.m.; taking melatonin as late as bedtime results in an entrained MO occurring no later than the afternoon. The correct clock time for administering melatonin to these individuals is wake time, which will result in an entrained MO at about 8–9 p.m. (Lewy et al. 2004b; Emens et al. 2006).

The above examples illustrate the importance of the melatonin PRC in the treatment of BFRs. The melatonin PRC is also important for treating the sighted: Entrained people should take melatonin about 5–6 hours before their DLMO10 (saliva DLMO3) for optimally causing a phase delay if it is given between CT 6 and CT 18, it causes phase advances; when it is given between CT 18 and CT 6, it causes phase delays. However, administration later in the afternoon will also cause robust phase advances. In order to achieve phase advances, melatonin is not given before wake time, so as to not interfere with sleep. However, an individual who awakens spontaneously in the middle of the night can take melatonin to cause a phase delay if it is given about CT 19 (1 a.m.). A sophisticated understanding of the melatonin PRC will be important when crafting melatonin treatment regimens and formulations. In addition to the human melatonin PRC, treatment recommendations are expected to be guided by the human melatonin dose-response curve (Fig. 4) (Lewy et al.
Relative Coordination to Weak Zeitgebers, the Tau Response Curve, and Tau at Entrainment Phase

In our first five BFRs assessed in a high-resolution protocol (MOs obtained as frequently as every 2 weeks throughout a complete circadian beat cycle [CBC], in which the MO traverses 24 hours), we observed a pattern typical of RC when the MOs were plotted longitudinally against clock time (Emens et al. 2005). Similar to free-running animals exposed to a daily zeitgeber too weak to effect entrainment, tau is longer than average in one half of the CBC and shorter than average in the other half. To precisely describe RC, we calculate the slope between each overlapping pair of successive MOs throughout one or more complete CBCs. The “two-point tau” is calculated as the slope plus 24 hours. The tau response curve (tauRC) is a plot of two-point taus against the midpoint between the clock times for the pair of MOs. For example, if the MO is at 8 a.m. on day 1 and at 10 a.m. on day 5, the tau between days 1 and 5 would be 24.5 hours. As the MO drifts later each day between about 8 a.m. and about 8 p.m., the two-point taus are longer than the average tau for each individual (the long-tau zone of the tauRC); as the MO drifts between 8 p.m. and 8 a.m., it is shorter than average (the short-tau zone of the tauRC), i.e., the crossover times of the tauRC are on average at about 8 a.m. and 8 p.m. Because the latter coincides with the average time of the DLMO in entrained sighted people, it appears that the weak zeitgebers are working to effect entrainment at the normal phase or a few hours later. BFRs with taus less than 24 hours have the same tauRCs; for them, the weak zeitgebers are perversely working to effect entrainment at an undesirable phase. Fortunately, there are relatively few such individuals (and they can be successfully treated with low-dose melatonin taken at wake time). Astronauts adapting to the Martian day of 24.65 hours are another group of individuals that might benefit from a small daily dose of melatonin taken at wake time (Lewy et al. 2004b).

In the dose-response study described above, a tauRC was calculated for each BFR, all of whom had taus greater than 24 hours and were administered melatonin about 6 p.m. in order to entrain them at the normal phase. The tauRC was used to calculate TEP. TEP is the two-point tau of an individual at the clock time of their now-entrained MO. This takes into account advances or delays working to effect entrainment at an undesirable phase. Fortunately, there are relatively few such individuals (and they can be successfully treated with low-dose melatonin taken at wake time). Astronauts adapting to the Martian day of 24.65 hours are another group of individuals that might benefit from a small daily dose of melatonin taken at wake time (Lewy et al. 2004b).

Average Tau and Human Tau Phenotypes

TEP, not average tau, was used to calculate the daily phase advance required for entrainment in the dose-response study, although when the MO is about 8 p.m., it is not much different from average tau. The standard method of calculating average tau is by linear regression through a longitudinal plot of the MOs. There is another way to calculate tau in BFRs, first reported by the Czeisler group, although there are errors in their algebraic equations (Klein et al. 1993). In any event, average tau can also...
be calculated by dividing 24 hours by the number of days of a complete CBC and then adding (for BFRs with taus greater than 24 hours) 24 hours. For example, if the MO is occurring at 2 p.m. on day 1 and then next occurs at 2 p.m. on day 48, average tau is 24.5 hours. Interestingly, these two ways of calculating average tau produce nearly identical results and may suggest that RC does not affect the average tau of an individual, another reason why the average tau in a blind person may be the most accurate estimate of intrinsic tau. The CBC method is most accurate when the beat cycle begins and ends (at the same time) in the long-tau zone. However, if any two-point taus overlap 24 hours (i.e., for most BFRs, become <24 hours), a confound will be introduced when calculating average tau (using any method). In most blind people, this “transient entrainment” would shorten tau, as would the effects (if any) of a 24-hour forcing period. Thus, the average tau in blind people of about 24.5 hours may be slightly less than the average of their “true” intrinsic taus, another reason why their taus compared to the shorter taus of sighted people (studied under FD) is such an important scientific issue (see below). Therefore, average tau should be corrected in the few BFRs who have transient entrainment (which can be detected with frequent phase assessments in the short-tau zone of the tauRC).

Aftereffects in sighted people resulting from entrainment to the 24-hour light/dark cycle (Sheer et al. 2007) have been offered as a possible explanation for their shorter average tau (most recently calculated to be 24.07 hours in 12 subjects studied in an FD T cycle of 28 hours under 1.5 lux during wakefulness; Gronfier et al. 2007) compared to blind people. However, aftereffects cannot explain why the range in taus in the sighted now appears to be the same as in the blind (about 1 hour) and why the proportion of people with taus less than 24 hours is now one-third (Gronfier et al. 2007), which is much lower in the blind (less than about 10%), in whom average tau is about 24.5 hours (Kendall and Sack 2000; Lockley et al. 1997; Suck et al. 1992b); however, definite statements require more sophisticated models about how exactly aftereffects are generated. Other explanations for the differences in tau between blind and sighted people have been sampling bias in the blind (excluding taus that are too close to 24 hours to be designated as free-running) or exposure of blind people to nonphotic time cues in field studies; however, sampling bias is probably not a sufficient explanation and the other explanation is unlikely (Kendall and Sack 2000; Emens et al. 2005).

The differences in average tau and the percentage of taus less than 24 hours between the blind and the sighted can be rectified by another explanation: Either ambient light of 1.5 lux FD shortens tau by up to 0.4 hour on average and/or becoming blind lengthens tau by up to 0.4 hour on average. In a thought experiment, we subtracted 0.4 hour from TEP (and from average tau) that we had calculated in our dose-response study (see Fig. 4) (Lewy et al. 2005). Even subtracting 0.2 hour produces a result contrary to what we found: Determining the baseline taus of these BFRs to be greater than 24 hours must have been correct, or they would not have entrained their MOs at the normal phase when taking exogenous melatonin in the early evening (BFRs with taus less than 24 hours entrain to an evening dose of melatonin with their MOs occurring in the afternoon). Therefore, it appears to be prudent to conclude that FD in dim light confounds the accurate measure of tau with an average shortening of 0.4 hour. This confound is probably consistent, given that these taus are apparently useful in correlational analyses (Wright et al. 2005; Gronfier et al. 2007); however, consistency remains to be documented. In any event, the study of tau in sighted people is problematic, given the recent reports of aftereffects (Sheer et al. 2007), which curiously appear to be different from those found in animals, in that the latter exhibit a gradual return to baseline tau which does not appear to be the case over several weeks of follow-up in humans. Studies (Wever 1979) using temporal isolation have been criticized (Czeisler et al. 1999) for introducing an artifact of lengthening tau of at least 0.4 hour. Now it appears that FD in dim light may be introducing a similar artifact in the other direction, perhaps the lower the light intensity, the greater the artifact (possibly a parametric effect of light). However, as light intensity increases above a certain threshold, it will tend to alter tau in the direction of the T cycle (i.e., $T = 28$ hours lengthens tau if light intensity is above a certain threshold).

The above discussion may help develop more accurate tau phenotyping in sighted people, which will be necessary for identifying the human clock genes that may be responsible for circadian phase disorders (for review, see Lamont et al. 2007). Similar to melatonin entrainment of blind people (see Fig. 3) (Lewy et al. 2001), PAE in sighted people entrained to the light/dark cycle was subsequently reported to correlate with tau under FD (Wright et al. 2005). When the issues discussed above, including those of aftereffects due to entrainment to the 24-hour light/dark cycle (as well as the best way to calculate PAE in sighted people), are settled, tau phenotyping can proceed apace and help in the identification of human clock genes and their disorders. Until then, molecular biologists can focus on those people with extremely short and extremely long taus, in whom the “true” intrinsic tau is less important. In the meantime, the measurement of tau in totally blind humans is closest to the established convention in which animals are studied under conditions of constant darkness. Perhaps the dichotomous difference between the very few BFRs with taus less than 24 hours and the vast majority who have taus greater than 24 hours will be particularly important in identifying the human clock genes.

It has also been suggested that altered innervation of the SCN from the retinohypothalamic tract might lengthen tau (Klerman 2001). However, animal studies do not support this (Bobbert and Riethoven 1991; Mistlberger 1991; Yamazaki et al. 2002). Furthermore, a change in tau correlating with the duration of blindness has not been observed cross-sectionally in either our data or those of others (Lockley et al. 1997). Whatever are the reason(s) for the discrepancy in average taus between sighted people studied under FD and blind people, their elucidation may lead to profoundly important advances in our understanding of human chronobiology. In any event, knowledge of a BFR’s tau—at least whether it is less than or greater than 24 hours (perhaps its most intrinsic character-
istic)—is critical when treating BFRs: For entrainment at the normal phase, melatonin should be administered in the early evening to most blind people; however, if a blind person’s tau is less than 24 hours, even just slightly less (i.e., 23.94 hours), tau must be administered in the morning for normal entrainment, which most conveniently would be wake time.

**Women May Be More Sensitive to Social Cues than Men**

Average tau is not needed for calculating the range of oscillation, or RC amplitude, which we define as the shortest two-point tau subtracted from the longest two-point tau in a BFR’s tauRC. RC amplitude varies greatly among blind people. Studies of this variability may help identify the as yet unknown weak zeitgebers responsible for the daily phase advances that appear to shorten observed two-point tauds and daily phase delays that appear to lengthen them (we do not believe that intrinsic tau is actually changing). In fact, RC amplitude may be a “bioassay” for sensitivity to these weak zeitgebers, particularly if they have consistent strength and timing.

Women appear to be more sensitive to the weak zeitgebers than men: Free-running blind women have twice the RC amplitude than men; comparing circadian status, using more rigorous definitions of entrainment and free-running than was customary (Sack et al. 1992b; Lockley et al. 1997; Klerman et al. 2002b; Lewy et al. 2003) before our report of RC in BFRs (Emens et al. 2005), we found no totally blind men who are naturally entrained, compared to 25% of women (Lewy et al. 2007c). The difference in circadian status proportion occurs after puberty; in fact, the proportion of naturally entrained prepubertal boys appears to be greater than that of girls. Circadian status may turn out to be another bioassay (in addition to RC amplitude) for social sensitivity, given that putative candidates for these zeitgebers are probably dependent in some way on social cues. These include sleep, activity, exercise, meal times, temperature, acoustic signals, and pheromones (Goel et al. 1998). Because we believe that the overwhelming majority of humans have taws greater than 24 hours, we hypothesized that sighted women would entrain at an earlier PAE. A search of the literature with this hypothesis in mind was fruitful (Campbell et al. 1989; Mongrain et al. 2004), as well as analysis of one of our own data sets (Emens et al. 2007). Interestingly, RC amplitude for our group of blind people (up to 56 minutes). A search of the literature with this hypothesis in mind was fruitful (Campbell et al. 1989; Mongrain et al. 2004), as well as analysis of one of our own data sets (Emens et al. 2007). Interestingly, RC amplitude for our group of blind people (up to 56 minutes).

**Circadian Disorders in Sighted People**

Knowledge of a sighted person’s tau is not needed for treating circadian disorders, which are mostly, if not exclusively, disorders of circadian phase position. In some cases, such as in ASPS and DSPS, the timing of sleep is sufficient for diagnosis and treatment. The first patient with DSPS treated with morning light was reported in 1983 (Lewy et al. 1983), and the first patient with ASPS treated with evening light was anecdotally reported in 1985 (Lewy et al. 1985b). Melatonin was first used to treat DSPS in 1990 (Dahlitz et al. 1991), although not at the optimal time to cause phase advances, according to the melatonin PRC (Lewy et al. 1998a). Aaron Lerner may be the first person reported to have been successfully treated for ASPS with the neurohormone he discovered (0.3 mg of melatonin at wake time) (Lewy 2007). Other circadian phase disorders amenable to melatonin (and light) treatment include jet lag (Arendt et al. 1987; Lewy et al. 1995; Burgess et al. 2003) and maladaptation to shift work (Sack et al. 1992a). Treatment has not progressed rapidly in the area of shift work, due to many factors, such as rapidly changing schedules and the fact that DLMO time cannot even be grossly estimated from sleep times in these individuals.

**The Phase-shift Hypothesis**

Perhaps the most complex circadian disorders are psychiatric. The precise mechanism for the circadian component of psychiatric disorders, as well as of poorly maintained and nonrestorative sleep, is likely based on internal circadian misalignment between the sleep/wake cycle (and rhythms related to it) and those circadian rhythms that are more tightly coupled to the endogenous pacemaker (Kripke et al. 1978; Wehr et al. 1979). Although melatonin is probably not mediating pathology, its circadian rhythm is the most useful for marking the phase of the endogenous circadian pacemaker that regulates circadian rhythms which may indeed be pathogenic if misaligned with the sleep/wake cycle. SAD, because of the predictable winter depressions and the antidepressant response to bright light, is an ideal model for understanding the role of circadian rhythms in psychiatric disorders, as are totally blind people for studying another unfortunate situation (indeed, experiment of nature) involving light deprivation.

According to our PSH, most patients with SAD become depressed in the winter at least in part because of a phase delay with respect to the sleep/wake cycle (Lewy et al. 1985c, 1987a,b, 1989) and that bright light should optimally be scheduled in the morning in order to provide a corrective phase advance. Following our publication that morning light is more antidepressant than evening light (Lewy et al. 1987a), a vigorous debate ensued with some (Avery et al. 1990) but not all (Wirz-Justice et al. 1993) investigators supporting the PSH. Superiority of morning light was finally resolved in 1998 (Eastman et al. 1998; Lewy et al. 1998b; Terman et al. 1998). This still did not establish the PSH, because it could be argued that there is an overall increase in light sensitivity in the morning. Even correlational studies between circadian phase and treatment response (for review, see Terman et al. 2001; Lewy et al. 2006b), although important, did not establish causality. This was done recently using melatonin, rather than light, to treat SAD patients (Lewy et al. 2006b). Our study appears to be the first report of psychiatric symptom severity correlating with a biological marker before and in the course of treatment in the same patients.
Unlike light, low-dose ("0.3 mg) melatonin cannot be detected by research subjects and has less of a placebo component. Furthermore, it would not be expected to be antidepressant in SAD unless used at the opposite time of the day (afternoon/evening) to cause phase advances similar to those caused by morning light. An early corollary of the PSH is that a smaller subgroup of SAD patients phase advance in the winter and require bright light in the evening that would provide a corrective phase delay; therefore, the correct time of administration is in the afternoon/evening for the prototypical phase-delayed patient and in the morning for the phase-advanced types. The phase-angle difference (PAD) between the DLM0 and mid-sleep assesses internal circadian alignment (Fig. 5).

Average PAD in healthy subjects, using the DLM010, is 6 hours (Lewy et al. 1998b). Accordingly, we hypothesized that depression ratings would be lowest in patients with PAD 6, which turned out to be the vertex of parabolic curves that significantly correlated depressive symptom scores plotted against PAD before and after treatment (subjects were randomly assigned to placebo, morning melatonin treatment or afternoon/melatonin treatment). Subjects with PAD 6 at baseline were designated as phase-delayed types, i.e., their DLM0s were delayed with respect to the sleep/wake cycle. Subjects with more than two PAD >6 were designated as phase-advanced types. More than two decades after proposing that subjects be phase typed with respect to the sleep/wake cycle, subjects with PAD >6 should preferentially respond to morning light in order to cause a phase advance, the greater the phase advance, the better. They may consider PAD to be an epiphenomenon. Indeed, they recommend using mid-sleep (or, alternatively, Horne–Ostberg morningness/eveningness ratings) to estimate the DLM0. They then recommend, even if it involves interfering with sleep, scheduling exposure to begin 8.5 hours after the estimated DLM0 in order to induce what we (Lewy and Sack 1989) and they (Terman et al. 2001) have found to be about a 1.5-hour phase advance. Although earlier light and earlier sleep times would cause greater phase advances in the DLM0, we advise not requiring SAD subjects to wake up earlier because this would tend to shorten, not lengthen, the DLM0/mid-sleep PAD (the Terman group only assessed the DLM0/sleep onset interval (MSI) as their marker for PAD). Another difference between the original PSH and the Terman variant is that the latter leads to a recommendation of the earliest morning light in whom they would consider to be their least phase-delayed patients, whereas we would consider at least some of the latter to be of the phase-advanced type and would recommend that light be scheduled in the evening in order to provide a corrective phase delay.

We consider the clock time of the DLM0, not PAD, to be an epiphenomenon: If sleep times are invariant, a change in PAD will result in a commensurate change in DLM0 clock time, but it is the change in PAD that is clinically important. In addition to representing internal circadian alignment, PAD is less likely than real (or estimated) DLM0 clock time to be confounded by changes in sleep times (either because of masking effects or because of an altered light/dark cycle) or to be accurately guessed by research subjects or their raters. Regarding the latter, if a person is phase typed as delayed because of a DLM0 that is delayed with respect to sleep, then sleep will be advanced with respect to the DLM0 in this individual, thus making it difficult to guess phase type correctly and thereby improving the double-blind integrity of the study.

**Figure 5.** The phase-angle difference (PAD) between the plasma DLM010 and the time of mid sleep is on average about 6 hours in healthy controls. PAD 6 can be used to phase type individuals and to assess internal circadian misalignment. (Adapted, with permission, from Lewy et al. 2006b [©National Academy of Sciences].)
Goodness of fit to PAD can also be used to validate clinical assessment and improve how best to refine them, as well as help to identify a clinical endophenotype. In an item-by-item analysis, we found that three items in the SIGH-SAD instrument could account for all of the statistically significant findings originally reported using the entire 29-item scale (Lewy et al. 2006b). These three items rated subjective depression severity, subjective anxiety severity, and amount of observed agitation at the interview (Lewy et al. 2007b). Therefore, it would not be surprising if mixed anxiety and depressive states had a circadian component that can be treated, probably adjunctively, with a phase-resetting agent. Even if the circadian component explains only 10% of the variance, melatonin—and to a lesser extent bright light—can be added to just about any treatment regimen with impunity. We are just beginning to appreciate the significance of correlating PAD with symptom severity in sleep, psychiatric and perhaps other medical disorders, as well as interindividual and intraindividual changes in cognition, mood, and attention in healthy controls.

CONCLUSION: THE FUNCTION OF ENDOGENOUS MELATONIN PRODUCTION IN HUMANS

Before concluding, we might speculate about the function of melatonin in humans. When we first described the melatonin PRC, we hypothesized that humans have retained the acute suppressant effect of light in order to facilitate the circadian phase-resetting effects of melatonin and to therefore augment entrainment of the endogenous circadian pacemaker by the phase-shifting effects of the light/dark cycle (Lewy et al. 1992). Our work with BFRs, however, now provides the basis for possibly a more important role to all sighted babies (Fig. 6). Exogenous melatonin entrains BFRs in part because the administration time is provided by the researcher or clinician. Similarly, the sighted mother sends her melatonin signal to the third trimester fetus (Reppert 1979) and to the sucking newborn (Ilnerová et al. 1993). Once a few months old, “circadian blindness” ends (Kleitman and Engelmann 1953; Tomioka and Tomioka 1991; McGraw et al. 1999), and the 24-hour light/dark cycle via the retinohypothalamic tract can entrain the endogenous circadian pacemaker.

Melatonin receptors in the SCN are functional by the third trimester (Reppert et al. 1988). Melatonin freely crosses the placenta (Reppert 1979) and is metabolized by the mother, so that both mother and fetus receive the same chemical signal for nighttime darkness. Whether or not melatonin levels in breast milk (which are about the same or are at least two-thirds those of plasma; Ilnerová et al. 1993) are sufficient for the entraining infant is not presently known. Some investigators have calculated the levels to be too low, as least for rat pups (Rowe and Kennaway 2002). However, with substantial activity of the catabolic enzyme converting melatonin to 6-hydroxymelatonin (Skene 2001) not occurring until after age 2–3 months (Sonnier and Cresteil 1998), sucking infants until this age may have a melatonin bioavailability closer to 100% than the adult’s 3–15%, as well as a smaller volume of distribution, resulting in a barely sufficient level for this function. Exposure of the mother to ambient light could reduce melatonin levels further. Thus, studies of BFRs may be applicable for synchronizing the sleep/wake cycles of perinates to their mother’s. Systems theory would suggest that, even if it makes no difference to the baby to be entrained or free-running, a less sleep-deprived mother will deliver better care. It may not be premature to

![Figure 6. Salivary pharmacokinetic (PK) profile and endogenous melatonin levels in a BFR entrained to a daily dose of 0.01 mg of exogenous melatonin. The dose is given at 18:00 (1 hour earlier than usual in order to better discern its PK profile), and the endogenous melatonin onset (MO) occurs a few hours later. In the steady-state situation, there is no melatonin-free interval between the exogenous and endogenous profiles, maximizing the phase-shifting effects of this very low dose, which must produce a daily phase advance of at least 0.1 hour to entrain this subject. Although saliva levels are less reliable than plasma levels (at least for higher concentrations, when the threefold difference between them becomes greater), there is no doubt that this dose is producing no more than a physiologic level for this subject. This finding suggests that very low doses may be therapeutic and that there may be a circadian function for endogenous melatonin production.](https://www.cs.cmu.edu/~symposium.cshlp.org/publications/lewy/figure6.png)
recommend that mothers avoid light exposure at night and to label pumped milk as to whether it was collected during the day or the night, so that the infant receives it at the appropriate time.

In conclusion, the DMLO can be used to phase type circadian disorders. The DMLO also marks the phase of the light and melatonin PRCs, informing treatment. Measurement of the DMLO in the course of treatment provides an objective measure of treatment efficacy and can indicate if too much of a phase shift has occurred. Thus, the DMLO provides information about whether or not a circadian phase disorder is present, its phase type, the appropriate choice and timing of treatment, and objective monitoring of treatment efficacy.

NOTE ADDED IN PROOF
A.J.L. is coinventor on several melatonin use-patents owned by Oregon Health & Science University currently not licensed to any company.

ACKNOWLEDGMENTS

We thank the research subjects and the nursing staff of the Oregon Health & Science University (OHSU) Clinical and Translational Research Center, Robert Sack, Jonathan Emens, Kyle Johnson, Mark Kinzie, Jeannie Songer, Krista Yuhas, Rebecca Bernert, Jennifer Rough, Neelam Sims, Anne Rutherford, Allie Witi, Diana Arntz, Cameron Brick, Cara Bussell, and Kitt Woods. This work was supported by Public Health Service grants R01 MH55703, R01 MH56874, R01 AG1826, and R01 HD42125 (to A.J.L.) and 5 M01 RR00334 (to the Clinical and Translational Research Center of OHSU). A.J.L. was supported by the National Alliance for Research on Schizophrenia and Depression 2000 Distinguished Investigator Award. J.S.E. was supported by Public Health Service grant K23 RR017636-01.

REFERENCES

Cajochen C., Jetew M.E., and Dijk D.J. 2003. Human circadian melatonin rhythm phase delay during a fixed sleep-wake schedule interspersed with nights of sleep deprivation. J. Pineal Res. 35: 149.


MELATONIN AND HUMAN CHRONOBIOLOGY

Chronobiol. Int. 19: 649.


Melatonin and Human Chronobiology

A. J. Lewy

Access the most recent version at doi:10.1101/sqb.2007.72.055

References
This article cites 130 articles, 15 of which can be accessed free at:
http://symposium.cshlp.org/content/72/623.full.html#ref-list-1

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.