Kaplan and colleagues report the findings of 2 studies that were part of a randomized clinical trial that aimed to determine whether a combined intervention for adolescents, using both light exposure during sleep and cognitive behavioral therapy, would encourage them to get to sleep earlier than usual, thereby increasing total sleep time. They conducted 2 similar, but somewhat different, studies.

In the first study, participants were randomly assigned to receive either 3 weeks of light flashes (light alone) or a sham light intervention. The light flashes were brief light pulses administered in 3-millisecond bursts delivered 20 seconds apart, starting 3 hours before the targeted wake time for the individual participant. In the second study, participants were randomized to a combination of light plus cognitive behavioral therapy or sham light therapy plus cognitive behavioral therapy. The main outcomes were self-reported sleep times, momentary ratings of evening sleepiness, and subjective measures of sleepiness and sleep quality.

Kaplan et al found that, in study 1, light therapy alone did not change sleep timing. However, in the second study, light plus behavioral therapy significantly moved sleep onset approximately 50 minutes earlier, on average, and increased nightly sleep time by approximately 43 minutes. There were also improvements in several secondary outcomes.

This combination of studies is similar in some respects to a factorial experiment but lacks certain unique advantages of factorial studies. The purpose of this commentary is to elaborate on those potential advantages of factorial studies, referring back to the article by Kaplan et al for context.

Factorial designs are used to test more than 1 experimental factor (whence the name) in the context of a single study. In the studies by Kaplan and colleagues, the 2 experimental factors were light and cognitive therapy. The first study addressed the efficacy of light in the absence of behavioral therapy. The second study tested the efficacy of light in the presence of behavioral therapy. The 2 studies were conducted sequentially. In a traditional factorial experiment, participants would have been randomized to 1 of 4 groups: light alone, cognitive therapy alone, both light and cognitive therapy, and neither. Another way to think of the design is that participants would be randomized to light vs sham; then, within each of those groups, they would be randomized again to cognitive behavioral therapy or no therapy.

The potential for statistical interaction, ie, when the effect of an intervention depends on the presence or absence of another intervention, is both a strength and a limitation. The strength is that the question of differential effects may actually be of scientific and clinical interest. The limitation is that if the interest of the investigators is solely on the effects of each intervention, assuming that there is no interaction, then the statistical power to address those individual questions is potentially compromised. Again, in the studies by Kaplan et al, the question is whether light flashes have different effects in the presence or absence of behavioral therapy.

In this situation, the investigators were uninterested in the effects of behavioral therapy alone. Nonetheless, the conducted studies confounded several design features that slightly complicate the interpretation of their findings. Just by virtue of doing separate studies, the opportunity for increased variability is introduced. The expectation in conducting randomized trials is that results can vary from study to study even when the designs are similar. The fact that the studies were conducted sequentially allows for further variability due to calendar time. It may be that neither of these sources of variability would be meaningful for these studies, but the design does not allow disaggregation of these extraneous features. The timing of the administration of light flashes (starting 3 hours before...
wake time in study 1 vs 2 hours before wake time in study 2) also changed. In this case, different findings about the efficacy of light therapy could be (and probably are) due to the presence of behavioral therapy, but it is also possible, especially given the small sample size in the second study, that unrelated study-to-study variability could have produced the findings (not to mention the timing of the administration of light flashes).

The use of factorial designs in medicine is not new. A very famous example was the Physicians' Health Study, the design of which is discussed in a methodological article. Their goal was to test 2 hypotheses in the same study: efficacy of aspirin in prevention of cardiovascular disease and of beta carotene in the prevention of cancer among US physicians.

For beta carotene, with follow-up over nearly 13 years, the trial found no significant decrease in cancer risk overall, or in risk of prostate, colon, or lung cancer. For myocardial infarction, there was a 44% reduction in risk (relative risk, 0.56; 95% CI, 0.45-0.70). There was no reduction in overall mortality (relative risk, 0.96; 95% CI, 0.60-1.54).

Given the potential advantages of factorial experiments, one might wonder why they are not used more frequently. As noted above, there can be reduced statistical power for testing the effects of each treatment when there is a statistical interaction between treatments characterized by 1 treatment (either presence or absence) reducing the effect of the other. An example of this is provided in the study by Jaki and Vasileiou. For any study with more than 1 treatment, there may also be logistical complexities introduced by the need for multiple placebos, with the associated burden on trial participants, who are asked to take multiple medications. My own view is that factorial studies are underused, for the reasons I have described. In particular, the ability to eliminate confounding between study, calendar time, and the 2 factors should be very appealing. Furthermore, if we are confident about the lack of interaction, the factorial design can be very efficient (because we have 2 studies done at the same time). When we are actually interested in testing the interaction, the factorial design offers a great opportunity. That said, Jaki and Vasileiou suggest the use of an alternative design that accommodates multiple treatments and note that these other designs can be equally, or more, statistically efficient. The theme is the same, though, that testing multiple treatments in the same study is advantageous compared with doing separate studies.

In summary, the authors' conclusions about the efficacy of light flashes are likely to be correct (within the limits of self-reported outcomes, which they note). That said, when designing studies that address more than 1 question, when possible, one should consider the potential advantages of unconfounding those factors by conducting a factorial experiment.