

PLEASE NOTE: This is the authors' copy of the accepted manuscript, which has not undergone copyediting. The copyedited version of this manuscript published in *SLEEP* is available at: <https://doi.org/10.1093/sleep/zsad199>

Please cite this manuscript as: Shaffer, K. M., Daniel, K. E., Frederick, C., Buysse, D. J., Morin, C. M., & Ritterband, L. M. (2023). Online Sleep Diaries: Considerations for System Development and Recommendations for Data Management. *Sleep*, zsad199.

Online Sleep Diaries:

Considerations for System Development and Recommendations for Data Management

Kelly M. Shaffer¹, Katharine E. Daniel¹, Christina Frederick¹, Daniel J. Buysse²,
Charles M. Morin³, & Lee M. Ritterband¹

¹ University of Virginia School of Medicine, Charlottesville, VA, USA

² University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

³ Laval University, Québec, QC, Canada

Corresponding Author: Kelly M. Shaffer, PhD
PO Box 801075
Charlottesville, VA 22908
Email: kshaffer [at] virginia [dot] edu

ABSTRACT

Study Objectives: To present development considerations for online sleep diary systems that result in robust, interpretable, and reliable data; furthermore, to describe data management procedures to address common data entry errors that occur despite those considerations.

Methods: The online sleep diary capture component of the Sleep Healthy Using the Internet (SHUTi) intervention has been designed to promote data integrity. Features include diary entry restrictions to limit retrospective bias, reminder prompts and data visualizations to support user engagement, and data validation checks to reduce data entry errors. Despite these features, data entry errors still occur. Data management procedures relying largely on programming syntax to minimize researcher effort and maximize reliability and replicability. Presumed data entry errors are identified where users are believed to have incorrectly selected a date or AM versus PM on the 12-hour clock. Following these corrections, diaries are identified that have unresolvable errors, like negative total sleep time.

Results: Using the example of one of our fully-powered, U.S. national SHUTi randomized controlled trials, we demonstrate the application of these procedures: of 45,598 total submitted diaries, 487 diaries (0.01%) required modification due to date and/or AM/PM errors and 27 diaries (<0.001%) were eliminated due to unresolvable errors.

Conclusion: To secure the most complete and valid data from online sleep diary systems, it is critical to consider the design of the data collection system and to develop replicable processes to manage data.

Keywords: Data analysis; Data collection; Internet; Sleep; User-Centered Design

Clinical Trial details: Sleep Healthy Using The Internet for Older Adult Sufferers of Insomnia and Sleeplessness (SHUTiOASIS); <https://clinicaltrials.gov/ct2/show/NCT03213132>;

ClinicalTrials.gov ID: NCT03213132

Statement of Significance: Online sleep diary systems offer many benefits relative to traditional paper-and-pencil methods. While there are clear guidelines regarding the content to be collected by sleep diaries, as well as processes to develop effective online diary user interfaces, there is no guidance regarding best practices for system back-end development or data management procedures. These considerations are necessary to promote the collection of valid and complete sleep diary data. To fill this gap, this manuscript presents features built into one online sleep diary collection system, refined over years of system testing, as well as data management processes used to address common data errors that occur through this system and may be pertinent however sleep diary data is obtained.

Introduction

Sleep diaries are a central component of diagnostic assessment, behavioral treatment, and outcome monitoring for insomnia disorder, as they track individuals' sleep patterns and behaviors across time. [1–3] While sleep diaries traditionally have been collected using paper-and-pencil methods, [3] it is now common to capture this data electronically, particularly given the rise of Internet-delivered insomnia treatment. [4] Online sleep diary systems hold multiple benefits over paper-and-pencil methodology for patients, clinicians, and researchers. These include reducing the backfilling of many days' worth of data at one time, thereby limiting memory errors; allowing clinicians to monitor patients' diaries via a shared database; standardizing data input; and reducing incorrect or problematic data with automated error checking. In general, online sleep diaries result in better quality data and facilitate data interpretation.

These benefits notwithstanding, challenges remain when leveraging electronic platforms to collect sleep diary data. For example, complications can arise from the date-time data structure, managing the volume of repeated assessments per user, and using appropriate and correct error checking algorithms to ensure useful and accurate data is entered. While there are clear guidelines regarding the content of sleep diaries (i.e., Consensus Sleep Diary [3]) and user-centered development processes for effective online sleep diary user interfaces (e.g., [5,6]), there is no published guidance regarding online sleep diary system development and data management practices. The current manuscript puts forward considerations and recommendations for such system development and data management practices for online sleep diary systems.

The Internet insomnia program called SHUTi (Sleep Healthy Using the Internet) includes an online sleep diary collection system. SHUTi is a fully-automated, interactive online

intervention that delivers the primary therapeutic mechanisms of cognitive-behavioral therapy for insomnia through six “Cores,” or lessons. Online sleep diaries (Figure 1) are used to both direct and tailor the intervention, as well as collect treatment outcomes (i.e., trial assessments). The content collected by the diaries is based on the Consensus Sleep Diary, [3] asking users to record their times to bed, trying to fall asleep, wake, and rise; duration of time needed to fall asleep; number of awakenings; duration of time awake after initially falling asleep; and sleep quality (see Figure 1 for item wordings). Additionally, diaries ask users to report certain behaviors from the prior day, including if they napped (and if so, for how long), consumed alcohol (and if so, time of last drink), or took medication to help with sleep (and if so, what medications). As part of the intervention, data from online sleep diaries are used to provide a tailored ‘sleep window’ (i.e., recommended total amount of time in bed, as well as bedtimes and rise times) as part of the sleep restriction therapeutic technique. As such, SHUTi users are required to complete a minimum of five sleep diaries within one week following the completion of the first Core to receive access to the second Core, which introduces sleep restriction and provides their initial personalized sleep window. Users continue to track their sleep using the online sleep diaries throughout the intervention, which allows the program to provide updated sleep windows over time.

SHUTi, including its incorporated online sleep diary system, has been tested in more than 25 trials across the world, with consistently strong participant-reported usability and satisfaction (e.g., [7–10]). As an example of the volume of online sleep diaries typically collected during a clinical trial, our fully-powered, U.S. national clinical trial of SHUTi among older adults accumulated more than 45,000 diary entries across 311 enrolled participants (SHUTiOASIS; 2018-2020; NCT03213132 – see Supplemental Section for additional details about these data).

Through two decades of research and hundreds of thousands of diaries, our team has learned key features of online sleep diary data collection systems that result in robust, interpretable, and reliable data.

Here, we first address development considerations of our online sleep diary system that are intended to promote data integrity. Second, we describe data management processes to address common data entry errors and calculate diary-based trial outcomes. This paper serves as a companion to the Sleep Diary Data Management Guide developed for projects undertaken through our “SHUTi for Researchers” platform (Sleep Diary Data Management Guide [11] openly available through GitHub at: <https://github.com/BHT/SHUTiForResearchers>). The “SHUTi for Researchers” platform is a multi-tenant instance of the SHUTi system made available to researchers for a fee (see <https://www.shuti.org>).

While the recommendations we make here are framed in relation to online sleep diaries, they could also be useful for many kinds of online self-reported symptom or behavior tracking. In addition, while the system development recommendations only pertain to electronic sleep diary data capture, our data management processes may be pertinent to sleep diary data regardless of how they are collected.

Considerations and Recommendations for the Development of Online Sleep Diary Systems

Comprehensive and iterative user testing is critical to designing an effective and engaging online sleep diary system; however, large-scale use in the real world is likely to reveal additional areas for improvement for even very well-tested systems. Moreover, while user testing refines the ‘front-end’ of the diary system (i.e., what users see and interact with), less consideration and testing are typically devoted towards the ‘back-end’ of the system (e.g., what data is collected and how it is stored). Back-end problems lead to difficulties with data export, analysis, and

interpretation, so it is essential to develop the back-end carefully to ensure data integrity.

Although it is likely impossible to prevent all data entry errors even with the most careful front-end and back-end design, we list recommendations and considerations below that will limit errors and support the collection of more complete and valid data (see Table 1 for summary).

Limit retrospective bias with diary entry restrictions. A major limitation of paper-and-pencil diaries is that people may ‘hoard’ their data, writing down their sleep data from all days of a week in one sitting. This so-called “parking lot syndrome” is problematic because of the length of time between the night’s sleep and the report about the night’s sleep, increasing recall bias and decreasing data reliability. [12] To prevent this from occurring, it has been previously recommended that online sleep diary systems limit how late a diary may be entered. [13] It is important to consider, though, how steps to minimize recall bias (e.g., requiring users to regularly log into a data capture system to record a diary) can exacerbate user burden and reduce diary completion rates, resulting in greater missingness or dropout.

In response to these competing considerations, SHUTi users are permitted to enter up to three consecutive sleep diaries (see Figure 2). For instance, assuming a user has not completed any of these diaries previously, on the morning of January 3, a user could complete last night’s diary (i.e., evening of January 2 into morning of January 3), yesterday’s diary (i.e., January 1 into January 2), and the previous day’s diary (i.e., December 31 into January 1). If that user still did not enter any diaries, on the morning of January 4, the user would no longer be able to enter a diary for December 31 into January 1. The three-day window was selected to balance data validity (i.e., limiting retrospective bias) with flexibility (i.e., greater accommodation for those unable or unwilling to log on every day).

Promote data entry by supporting engagement. Unlike paper-and-pencil diaries, online sleep diary systems can both monitor sleep diary engagement and directly motivate engagement through the diaries themselves. The term “engagement” is used here to describe an individual’s continued submission of sleep diaries. While the term “adherence” is also often used, “engagement” better acknowledges the interactive nature and importance of a sleep diary system itself in sustaining the user’s data entry across time. Theoretical models and frameworks can help guide the selection of strategies to enhance engagement with online diary systems, such as the Model for Internet Interventions [14] and the gamification framework [15].

One strategy to support engagement is to provide reminders, or prompts, to complete diaries. Automated reminders – whether by email, text messages, or app push notifications – are valuable, low-resource tools that keep users engaged in health behaviors. [4,16] In fact, user feedback testing suggests users want reminders to support their engagement with diary completion, [6] and email reminders have been successful in encouraging users to engage with Internet interventions. [17] The SHUTi sleep diary system automatically sends daily email reminders to participants to complete their sleep diaries.

Another strategy to support engagement with online sleep diaries is to return valuable insights to users, gained from their data, to motivate users to complete additional diaries. Receiving personal data-driven insights like reports or data visualizations can provide significant motivation for users to remain engaged. For example, users have described data visualization components of online sleep diary systems as rewarding and helpful for enhancing self-knowledge, thus reinforcing their engagement. [5,6] SHUTi incorporates such data visualization, providing users with graphs of their sleep metrics across time (Figure 3). Specifically, users see a bar graph depicting their reported total amount of time they slept and reported total amount of

time they were awake each night. This graph also visualizes their nightly “sleep efficiency,” or the percentage of time that they reported being asleep relative to the amount of time they reported being in bed, with a goal to achieve efficiency scores of 85% or above. Users are also presented a bar graph depicting the number of times they reported waking during the night and the reported total time of these wake periods. Finally, users can also track their reported time to bed, time to fall asleep, time to wake, time out of bed, medication use, and subjective sleep quality for each day, all displayed in a chart. This graphic summary of users’ data holds the double benefit of both encouraging user engagement and reinforcing progress towards treatment goals.

In addition, the SHUTi intervention also provides value in return for regular diary completion by using diary data to tailor ‘sleep windows’ (i.e., bed time and rise time) as part of sleep restriction. Users receive updated sleep windows over the course of the intervention if they continue entering sleep diaries. Users are typically motivated to do so, given that prescribed time in bed (and therefore likely total sleep time) is increased as their sleep improves.

Promote accurate data entry through the diary presentation. Time-date data can be confusing for users to enter. User confusion can be minimized through user-centered design approaches, which guide product development based on the preferences and needs of those that the product is intended to serve. [18] This process of clearly identifying the intended product users and their needs, sharing prototypes with intended users, and making changes based on user feedback can ultimately make a product more appealing, usable, and effective. [19]

Consistent with this approach, our team has conducted rigorous, iterative, and routine user testing on the SHUTi sleep diaries for a seamless user experience that encourages users to enter accurate data. For example, to prepare for the SHUTiOASIS trial among adults aged 55

and older, we completed a series of one-on-one interviews with previous SHUTi trial participants who had experience with the sleep diaries. We specifically identified past participants who were over age 55 and divided interviews between users who reported complaints about the diary during their trial participation versus no complaints. These interviews generated ideas for refining the interface for a more seamless experience and to be less error-prone, for example, by eliminating drop-down menus. Insights were then integrated into different prototypes (e.g., replacing existing drop-down menus with time entry analog clock face, slider bar, or point-and-click), and we enlisted older adults with no prior experience with the SHUTi diary to interact with these prototypes using a “think-aloud” approach. [20] This second-round feedback led to our implementation of features, including a point-and-click button interface to input time, as well as presenting all of the sleep diary questions on a single page with only one item active at a time. These improvements have been implemented into our current sleep diary interface across studies and in the SHUTi for Researcher system.

These user insights, as well as experience across multiple large-scale trials, has helped us to refine the diary to make data entry as simple and clear as possible. Users select a diary to enter by clicking on an available diary date (see Figure 2). At the top of each diary, there is a depiction of the dates for the evening and morning to orient users. Due to user testing preferences described above, all of the sleep diary questions are presented queued and isolated, with items highlighted individually and remaining questions ‘greyed out’ (Figure 4). This allows users to focus on the present question, while also seeing what questions come next. Users begin by selecting the time that they went to bed as hour, minute (in 5-minute intervals), and AM/PM using a simple point-and-click or touch screen interface. Then, system logic infers the date (e.g., for a January 2 diary, if a user entered 11:00 PM, the system would infer a bedtime date of

January 1), although users are able to override the determination should it be incorrect. For each diary, only two dates are permitted – the day of the diary (e.g., January 2) and the day prior to the diary (e.g., January 1). The same process is repeated for the sleep, wake, and rise times. Diary webpages are optimized for mobile devices, so diaries may be entered on any Internet-enabled computer, tablet, or smartphone.

It is important to accommodate local standards for reporting dates and times. The SHUTi program, for instance, was initially developed in the U.S., where dates are written in the order of Month-Day-Year and the 12-hour clock is utilized. We therefore denote months with written abbreviations (e.g., Jan 02 2023) rather than numerals (e.g., 01/02/2023) to limit confusion among users in areas with a Day-Month-Year date convention. Similarly, when SHUTi was translated for use in Norway, the 24-hour clock was required to reflect local standards. Having a system that can toggle between the 12-hour and 24-hour clock is ideal, but this requires planning and careful consideration in both how the data is stored and displayed.

Instructions for completing each item of the sleep diary, as well as how to enter more complicated sleep situations, should be readily available. In SHUTi, diaries have ‘just-in-time’ support through hyperlinks on each question that reveal pop-up windows of help instructions and example data entry (i.e., see Figure 1 “(Show More)” hyperlinks). An FAQ page provides specific instructions and examples for situations resulting in potentially confusing diary data entry, such as waking in the middle of the night and not falling back to sleep.

Promote data accuracy with restrictions. Incorporating robust data validation checks into diary entry systems minimizes missing or errant data entry. In the SHUTi system, there are several layers of automated data validation. Answers to each of the questions related to sleep outcomes are required (prior day behavior questions are optional), so there is no missing data.

The SHUTi system will display errors if the user attempts to submit impossible data and will not permit the diary to be submitted until data is corrected. Specifically, times must be chronological: time to bed must be at or before time trying to fall asleep, which must be at or before time awakening, which must be at or before time of rising. For instance, if a user indicates that they got into bed at 11:30 PM and began trying to fall asleep at 11:15 PM, a warning message would state, “The time you got into bed (answer to Question 1) can’t be later than the time you tried to go to sleep (answer to Question 2). Check your answers to Questions 1 & 2.” In addition, the SHUTi system also requires that the total minutes of sleep time – calculated as time of final awakening minus time of trying to fall asleep, minus minutes of sleep onset latency and wake after sleep onset – must be greater than or equal to 0.

Users are also encouraged to review their own data for accuracy. The SHUTi system accomplishes this through presenting users a “Sleep Story.” Once the user has completed a diary, users are shown their data on the “My Sleep Story” screen (see Figure 5). The Sleep Story feeds back key metrics to the user in a more conversational manner (e.g., “It took me 20 minutes to fall asleep.”), serving as an opportunity for users to check their data. For instance, a user who erroneously entered that they went to bed at 12:00 AM on January 2 and rose from bed at 7:00 PM on January 2 may realize their error on review. From the My Sleep Story screen, users are able to either return to their sleep diary to correct their entry or submit their diary if data appear correct.

It is also worth considering whether a user should be permitted to edit their data after submitting a diary. In SHUTi, once a user has submitted their diary following the My Sleep Story screen, that diary is locked and can no longer be edited by the user. If a user later notices that they entered incorrect data (for example, after viewing their sleep data graphs [Figure 3]),

staff can be contacted to correct erroneous diary data on their behalf. This process was implemented to prevent users from going back and changing their data to influence the system-driven tailoring (e.g., provide a wider ‘sleep window’).

Support data management with back-end data capture and formatting. Time-date data are complicated; time data that crosses calendar dates are even more complicated and often not intuitively so. To facilitate data cleaning, interpretation, and analysis, it is worth carefully considering what variables need to be assessed from users, how those data should be captured and formatted, and what data should be captured automatically by the system.

In the SHUTi system, in addition to the actual data being reported by users, diaries are automatically tagged by the system with several key pieces of information: ‘Diary ID’ – a unique number identifier for every submitted diary; ‘Participant ID’ – the unique identifier for each individual user; ‘calendar date’ – the date of the dawning morning of the diary; ‘diary type’ – whether the diary is part of the assessment or intervention periods; and, for intervention diaries, ‘Core’ to indicate with which intervention portion the diary is associated (for instance, diaries completed following the completion of Core 1 and leading up to Core 2 are tagged “Core 1”). In addition, diaries are automatically stamped with the time and date of their submission. These data settings require no additional effort by the user.

Users’ self-reported time-date data (e.g., time and date to bed) are recorded in two formats, which both help facilitate the data management processes described below. First, the data are logged in their local time-date (users indicate their time zone prior to completing their first sleep diary) on the 24-hour clock. Data are also represented with the Coordinated Universal Time (UTC), or the time at 0 degrees longitude without adjustments for daylight saving time, which is the formatting used to compute all outcomes metrics. Daylight saving time, or the

practice of advancing clocks so that sunset occurs later in the day, and the subsequent return to standard time is practiced across most parts of Canada, the United States, New Zealand, and Europe, as well as some parts of Australia and South America. Start and end dates, as well as the time of changes, vary across countries. Using UTC time-date data formatting obviates the need to account for these changes.

Recommendations for the Management of Online Sleep Diary Data

Attention to designing the front-end and back-end of online sleep diary systems for data integrity is important but unlikely to prevent all possible data entry errors. A typical SHUTi clinical trial collects tens of thousands of diaries, so mistakes will inevitably occur; however, this also makes it infeasible to examine and rectify each diary individually. Data management processes can help mitigate common sleep diary data entry errors in a scalable way. For researchers using the SHUTi system, processes to address common problems are operationalized in the R Markdown [21] Sleep Diary Data Management Guide [11] (file openly available through GitHub at: <https://github.com/BHT/SHUTiForResearchers>). Table 2 summarizes these recommendations, and Table 3 details these procedures with examples and numbers of diaries affected from our SHUTiOASIS trial. The following sections describe these common errors and the logic behind our responses to these errors, which should be applicable across sleep diary data capture systems.

Fixing date errors. As described above, users sometimes confuse the date for bedtimes and sleep times, even when provided a restricted range of entry options and the opportunity to review their data. In the SHUTi system, while date errors may technically occur for bed, sleep, rise, and wake times, such errors for rise and wake times should be extraordinarily rare and have not previously been detected in our trials (given our trial eligibility criteria requiring traditional

wake-sleep schedules, see Limitations sections for discussion). This is because it is extremely rare for an individual to have a final waking time between 6:00 PM and midnight, which could lead to a date error. For this reason, only bed and sleep times are evaluated for date entry errors.

Our data management procedures presume that a diary reporting more than 24 hours in bed or asleep, AND with a bed or sleep time from 12:00 AM – 3:00 AM (i.e., 00:00 to 03:00 on 24-hour clock), has an incorrectly entered date. To correct this particular error, 24 hours are added to the presumed erroneous bed and/or sleep time(s). Consider the example of a user who input a bedtime of 1:00 AM on January 1, sleep time at 2:00 AM on January 2, with a wake and rise time of 6:00 AM on January 2. For such a diary, 24 hours would be added to the presumed erroneous bedtime. Corrected data would reflect a bedtime of 1:00 AM on January 2.

We selected our time-based cut points (e.g., 3:00 AM as specified above) to identify diaries with likely erroneous dates (or AM/PM selection, described below) based on patterns observed from our review of hundreds of problematic diaries across our trials. These cut points best balance identifying where there is a strong reason to believe data are invalid due to user error versus otherwise unusual data. Consider the example of the following diary (and compare it to the example listed in the paragraph above): a user enters a bedtime of 5:00 PM on January 1, sleep time of 11:00 PM on January 1, wake time of 10:00 AM on January 2, and rise time of 6:00 PM on January 2. This diary is outside the norm, yet not inconceivable – perhaps the user was bedridden from injury or illness. Setting data management processes with criteria that can be justified and are acceptable to the research team allows for more replicable, reliable, and feasible data processing than *ad hoc* review.

Fixing AM/PM errors. Recording time on the 12-hour clock introduces the additional risk of entering an incorrect time by mis-selecting AM versus PM (e.g., indicating bedtime at

11:00 AM instead of PM). We have observed particular user confusion regarding midnight. Correctly entered, assuming a typical night's sleep, midnight is 12:00 AM on the date of rising (00:00 on the 24-hour clock). Imagining a user's true bedtime was midnight on January 2 (i.e., 12:00 AM on January 2 – for example, with a rise time of 7:00 AM on January 2), common user errors have been to enter 12:00 PM on January 1, or else 12:00 AM on January 1.

AM/PM selection errors can occur for bed, sleep, wake, and rise times. For wake and rise times, we manage data with the following presumption: a diary reporting more than 12 hours asleep or in bed, AND with a wake or rise time between 3:00 PM – 11:59 PM (i.e., 15:00 – 23:59), is considered to have an incorrectly entered PM instead of AM. Consider the example of a user who input a bed time of 10:00 PM on January 1, sleep time of 10:30 PM on January 1, and reported waking and rising at 5:00 PM on January 2. For such a diary, 12 hours are subtracted from the presumed erroneous wake and rising times. Corrected data would reflect waking and rising times of 5:00 AM on January 2.

For bed and sleep times, we manage data with the following presumption: a diary reporting more than 12 hours asleep or in bed, and with a bed or sleep time between 5:00 AM – 3:00 PM (i.e., 05:00 – 15:00), is considered to have an incorrectly entered PM instead of AM. Consider the example of a user who input a bed and sleep time of 1:00 PM on January 1 with wake- and rise times of 6:00 AM on January 2. For such a diary, 12 hours are added to the presumed erroneous bed and sleep times. Corrected data would reflect bed and sleep times of 1:00 AM on January 2.

It is important to note that well-constructed sleep diary systems should result in limited errors requiring correction. For context, in our SHUTiOASIS trial with 311 older adult

participants, only 487 of a total 45,598 online sleep diaries (0.01%) required modification by the above processes for fixing presumed errors in date and/or AM/PM data entry.

Identifying and removing diaries with unresolvable data entry errors. Fixing the date and AM/PM errors can introduce new errors in resulting diaries that cannot be clearly resolved. A diary with an unresolvable data entry error would not pass built-in system validation checks described earlier if the user attempted to enter the modified data, but the presumably erroneous data allows the diary to be submitted. One way this occurs is that times become out of order once fixed. For example, a user enters a bed time of 10:30 AM on January 1, sleep time of 10:15 PM on January 1, and wake and rise times of 7:00 AM on January 2. This diary would be corrected with our data management processes to have a bedtime of 10:30 PM on January 1, while still showing a sleep time of 10:15 PM on January 1. While it may be argued that the individual meant to swap the times (i.e., 10:15 PM bed time; 10:30 PM sleep time), our research team has conservatively decided to not make these additional assumptions. Rather, the entry is considered to have an unresolvable data entry error, and the diary is eliminated from the analyzed dataset.

Another way a diary can end up with an unresolvable error is when data cleaning processes result in a diary having less than 0 minutes of sleep time or time in bed. This can occur once a corrected total sleep time is computed from fixed sleep and wake times, factoring in self-reported minutes of sleep onset latency (SOL) and wake after sleep onset (WASO). Consider the example of a user who (presumably) erroneously input bed and sleep times of 2:00 AM January 1, wake and rise times of 5:00 AM on January 2, 1 hour of SOL, and 2.5 hours of WASO. Data management processes would result in this user's corrected bed and sleep times as 2:00 AM on January 2, for an updated total possible sleep time and time in bed of 3 hours. However, this user also reported 3.5 hours of time in bed awake (SOL + WASO), which would result in an

impossible negative 30 minutes of total sleep time. Such a diary is considered uninterpretable by our research team and is eliminated from the analyzed dataset. These types of errors tend to be rare. For context, in the SHUTiOASIS trial, 27 of a total 45,598 online sleep diaries (<0.001%) were eliminated from the analyzable dataset due to unresolvable data entry errors.

Additional review of atypical diaries. The processes outlined above address the majority of problematic diaries; however, we also recommend a comprehensive review of remaining diaries with unusual characteristics to determine if there are any other systematic problems that should be addressed across diaries. For example, we recommend reviewing diaries with 8 or more hours between reported bed time and sleep time (e.g., a user reports getting into bed at 2:00 PM and trying to fall asleep at 10:00 PM), 14 or more hours of total time in bed (i.e., bed time to rise time), and 14 or more hours of sleep opportunity (i.e., sleep time to rise time). Further, reviewing text entries in a “comments” or “notes” section, if available, can help determine whether individual diaries may or may not be valid.

The SHUTi data management syntax also identifies diaries that meet these different criteria and isolates them in a separate dataframe for researchers to review. This step is unlikely to place significant additional burden on the research team as the rate of occurrence of such diaries is typically small. For context, in the SHUTiOASIS trial, only 32 of the 45,571 valid sleep diaries (<0.001%) met any of these additional data review criteria. None of these identified diaries were ultimately modified or eliminated from the analyzable data set (i.e., all deemed valid). At the conclusion of this step, the result should be that all remaining diaries have acceptable validity and can be included in an analyzable (and interpretable) dataset.

Selecting diaries for trial assessment outcomes. When online sleep diaries are used as part of clinical trial outcomes assessment, the final step of data management is to select the

diaries that will be used to compute summary statistics. For assessment purposes in SHUTi trials, sleep diaries are typically requested from a participant for 1 to 2 weeks at each time point. This duration allows for more robust findings, parsing out normal night-to-night variation, without undue burden. It is also important to consider the proximity of diaries to the index event. In the case of a clinical trial, an index event might be the completion of a post-assessment battery of self-reported questionnaires or the completion of the intervention.

We use the following guidelines in our SHUTi trials to ensure robust assessment outcomes from sleep diaries that reflect the intended assessment period. Beginning at the baseline assessment, we require participants to complete their sleep diaries within 30 days of being entered into the SHUTi system (which may occur at time of consent or following completion of baseline questionnaires, depending on study design). If a participant does not do so, they are withdrawn from the study. Once 10 diaries have been entered within a 14-day period (requirement detailed below), the system automatically “unlocks” the participant’s allotted condition (e.g., SHUTi, online education control), affording them immediate access to their program.

At follow-up assessments, we require that assessment diaries are completed within 60 days of the assessment diary period opening. For example, our post-assessment is traditionally administered following the 9 week intervention period (i.e., Day 63) from the first day SHUTi is made available to a participant (i.e., Day 0). Therefore, the window for post-assessment diaries is from Day 63 to Day 123 (i.e., $63 + 60$ days). If a participant does not complete diaries within this period, they are considered lost to follow-up at this time point and approached again at the next assessment diary period. These cutoffs help ensure that diaries selected for a certain assessment

time point are indeed reflective of that assessment time point, with a longer period afforded at follow-ups to support retention for the longitudinal data collection.

As previously mentioned, we also require assessment diaries to be completed within a 14-day block. This ensures that diaries reflect a two-week window and guards against participants cherry-picking certain nights to report (e.g., only logging ‘good’ nights). Additionally, outcomes are ideally summarized from 10 sleep diaries, but can be summarized from a minimum of 7 diaries within the 14-day window. This helps ensure that the most robust and reliable summary statistics can be drawn from the block of diaries.

Combining these guidelines, we specify a set of rules to prevent selectively reporting certain nights for assessment diaries and to enforce consistency. Namely, we select a block of participant assessment diaries in the following order from within the 30- or 60-day assessment diary period: (1) first block of 10 diaries within 14 days (e.g., diaries dated according to rise date of January 1 through January 15); (2) if no block of 10 diaries, first block of 9 diaries within 14 days; (3) if no block of 9 diaries, first block of 8 diaries within 14 days; (4) if no block of 8 diaries, first block of 7 diaries within 14 days; and (5) if no block of 7 diaries within 14 days, then no outcomes will be computed for this participant at this time point.

For context, in the SHUTiOASIS trial with 311 participants, the study design included 4 assessment time points resulting in a total possible 1244 assessment diary periods (i.e., 311 participants x 4 assessment periods = 1244 total possible assessment periods). Accounting for withdrawals and loss-to-follow-up, 1110 assessment diary periods were initiated by participants. Of these 1110 assessment diary periods, outcomes were computed from participants completing 10 assessment diaries for 1071 assessment diary periods (96.49%), from 9 diaries for 11 assessment diary periods (0.99%), and from 8 diaries at only 1 assessment diary period (0.09%).

No assessment diary periods needed to have outcomes computed from a block of only 7 diaries (0%). There were 27 assessment diary periods (2.43%) where a participant initiated sleep diaries but had fewer than 7 valid diaries within a 14 day period in the assessment diary window, so diary outcomes were not computed for participants at these assessment periods (see Figure 6).

Discussion

Limitations. We continue to refine our system to reduce errors and improve data quality, but data entry errors will almost certainly always be possible and require data management. Certain circumstances forbid any post-hoc data changes described in our data management recommendations – for example, when submitting data for prescription digital therapeutic approval by the U.S. Food and Drug Administration (FDA). When such circumstances might apply, design considerations for the collection system will be even more critical to ensure the highest quality data. It should also be noted that our data management procedures have been developed based on our trial experiences. SHUTi trials have largely required individuals to report traditional sleep-wake hours for inclusion (i.e., usual bedtimes between 8:00 PM – 2:00 AM and arising times between 4:00 AM – 10:00 AM) and excluded individuals with regular or rotating night shift work schedules. Thus, our data management recommendations may not be appropriate or sufficient in samples where these criteria do not apply. In addition, trials have primarily included individuals residing in the U.S., Australia, or western European countries, reflecting the customs of these regions – certain recommendations may not be appropriate for all cultures, particularly where sleeping customs differ.

Conclusions. Sleep diaries are a key tool for assessing sleep, diagnosing sleep disorders, and directing behavioral sleep treatment. Online sleep diaries overcome many of the limitations of traditional paper-and-pencil diaries by better supporting engagement and mitigating

retrospective bias. Over two decades of testing and hundreds of thousands of diaries entered into the SHUTi online sleep diary system, our research team has learned much regarding the development of such systems, as well as analysis of their data. We hope these insights are helpful to all who conduct research with online diary data systems or collect sleep diary data. To secure the most complete and valid sleep diary data through online platforms, thoughtful and thorough user-centered design is necessary but not sufficient: considerations about the data collection system design and replicable processes to manage data are also necessary.

Acknowledgement: Authors thank Fabian Camacho, Holly Lord, and Frances Thorndike for their contributions to the development of our sleep diary data management processes. Authors also thank the software engineers responsible for designing, implementing, and maintaining our SHUTi sleep diary system: Rom DuPlain, Gabe Heath, Steve Johnson, Alan Lattimore, Nicole Le, and Ian Terrell.

Data Availability Statement: De-identified row-level data presented in this article are available via data use agreement with the University of Virginia per institutional requirement; contact the corresponding author to initiate the contracting process. Data management syntax is available as part of our open-access SHUTi for Researchers Sleep Diary Data Management Guide at:

<https://github.com/BHT/SHUTiForResearchers>

Financial Disclosure Statement: Collection of data presented in this manuscript was supported by the National Institutes of Health (R01AG047885; PI: Ritterband). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Shaffer, Daniel, and Frederick have no financial conflicts of interest to report. Buysse is an author of the Pittsburgh Sleep Quality Index, Pittsburgh Sleep Quality Index Addendum for posttraumatic stress disorder, Brief Pittsburgh Sleep Quality Index, Daytime Insomnia Symptoms Scale, Pittsburgh Sleep Diary, Insomnia Symptom Questionnaire, and RU_SATED (copyrights held by the University of Pittsburgh). These instruments have been licensed to commercial entities for fees. He is also a coauthor of the Consensus Sleep Diary (copyright held by Ryerson University), which is licensed to commercial entities for a fee. Morin has served on

advisory boards for Eisai, Merck, and Phillips; received research support from Idorsia and Canopy Health; and owns equity in BeHealth Solutions. Ritterband reports relationships with BeHealth Solutions (equity ownership and member of the board of directors), a company that was established to develop and disseminate digital therapeutics. Ritterband is also a consultant to Mahana Therapeutics, a separate digital therapeutic company not affiliated with this research. These companies had no role in the preparation of this manuscript. The terms of these arrangements have been reviewed and approved by the University of Virginia in accordance with its policies.

Non-Financial Disclosure Statement: None.

REFERENCES

1. Morin CM, Espie CA. *Insomnia: A Clinical Guide to Assessment and Treatment*. Kluwer Academic/Plenum Publishers; 2003.
2. Buysse DJ, Ancoli-Israel S, Edinger JD, Lichstein KL, Morin CM. Recommendations for a standard research assessment of insomnia. *Sleep*. 2006;29:1155-1173.
3. Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: Standardizing prospective sleep self-monitoring. *Sleep*. 2012;35(2):287-302.
4. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia: A systematic review and meta-analysis of randomized controlled trials. *Sleep Med Rev*. 2016;30(1):1-10. doi:<https://doi.org/10.1016/j.smrv.2015.10.004>
5. Blake J, Kerr D. Development of an online sleep diary for physician and patient use. *Knowl Manag E-Learn*. 2010;2(2):188.
6. Schmitz L, Sveinbjarnarson BF, Gunnarsson GN, et al. Towards a Digital Sleep Diary Standard. Published online 2022.
7. Thorndike FP, Saylor DK, Bailey ET, Gonder-Frederick L, Morin CM, Ritterband LM. Development and perceived utility and impact of an Internet intervention for insomnia. *E-J Appl Psychol Clin Soc Issues*. 2008;4(2):32-42.
8. Luyster FS, Ritterband LM, Sereika SM, Buysse DJ, Wenzel SE, Strollo PJ. Internet-based Cognitive-Behavioral Therapy for Insomnia in Adults with Asthma: A pilot study. *Behav Sleep Med*. 2020;18(1):10-22. doi:10.1080/15402002.2018.1518229
9. Ritterband LM, Bailey ET, Thorndike FP, Lord HR, Farrell-Carnahan L, Baum LD. Initial evaluation of an Internet intervention to improve the sleep of cancer survivors with insomnia. *Psychooncology*. 2012;21(7):695-705. doi:10.1002/pon.1969
10. Hagatun S, Vedaa Ø, Nordgreen T, et al. The Short-Term Efficacy of an Unguided Internet-Based Cognitive-Behavioral Therapy for Insomnia: A Randomized Controlled Trial With a Six-Month Nonrandomized Follow-Up. *Behav Sleep Med*. 2019;17(2):137-155. doi:10.1080/15402002.2017.1301941
11. Shaffer KM, Daniel KE, Ritterband LM. Sleep Diary Data - Formatting and Outcomes Computation for SHUTi for Researchers (Version 2.0) [Source code]. doi:10.5281/zenodo.7938322
12. Gorin AA, Stone AA. Recall biases and cognitive errors in retrospective self-reports: A call for momentary assessments. In: *Handbook of Health Psychology*. Lawrence Erlbaum; 2001:405-413.

13. Arnardottir ES, Islind AS, Óskarsdóttir M. The future of sleep measurements: a review and perspective. *Sleep Med Clin*. 2021;16(3):447-464.
14. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for Internet interventions. *Ann Behav Med*. 2009;38(1):18-27. doi:10.1007/s12160-009-9133-4
15. Floryan MR, Ritterband LM, Chow PI. Principles of gamification for Internet interventions. *Transl Behav Med*. Published online April 4, 2019. doi:10.1093/tbm/ibz041
16. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. *Soc Sci Med* 1982. 2013;97:41-48. doi:10.1016/j.socscimed.2013.08.003
17. Ritterband LM, Borowitz SM, Cox DJ, et al. Using the internet to provide information prescriptions. *Pediatrics*. 2005;116(5):e643-e647. doi:10.1542/peds.2005-0404
18. Baxter K, Courage C, Caine K. *Understanding Your Users: A Practical Guide to User Research Methods*. Morgan Kaufmann; 2015.
19. Lyon AR, Koerner K. User-centered design for psychosocial intervention development and implementation. *Clin Psychol Sci Pract*. 2016;23(2):180-200.
20. Jaspers MWM, Steen T, Bos C van den, Geenen M. The think aloud method: a guide to user interface design. *Int J Med Inf*. 2004;73(11):781-795. doi:10.1016/j.ijmedinf.2004.08.003
21. R Core Team. R: A language and environment for statistical computing [Computer software]. Published online 2021.

Online Sleep Diaries: Considerations for System Development and Recommendations for Data Management

System Development Recommendations

- Limit retrospective bias
- Promote data entry by supporting user engagement
- Promote accurate data entry through diary presentation
- Promote data accuracy with restrictions
- Support data management with back-end data capture and formatting

Data Management Recommendations

- Address common & systematic time-date errors
- Review remaining atypical diaries
- Select diaries from the validated set for outcomes computation

*R syntax for above processes
available at:*

<https://github.com/BHT/SHUTiForResearchers>




Diaries	You have added 0 diaries for the assessment. Show More					
	Sat Jan 7	Sun Jan 8	Mon Jan 9	Tue Jan 10	Wed Jan 11 Add >	Thu Jan 12 Add >
					Fri Jan 13 Add >	

My Sleep Diary Entry For




For the most complete and valid data from online sleep diaries,
consider the system design and develop replicable processes to manage diary data.

Figure 1: SHUTi Online Sleep Diary Data Entry Screen


Help Back

My Sleep Diary Entry For

Thu Jan 12



Fri Jan 13

- What time did you get into bed? [\(Show More\)](#)

10:00 PM Thu Jan 12 2023
- What time did you try to go to sleep? [\(Show More\)](#)

10:30 PM Thu Jan 12 2023
- How long did it take you to fall asleep? [\(Show More\)](#)

20 minutes
- How many times did you wake up, not counting your final awakening? [\(Show More\)](#)

☐ I didn't wakeup

2
- In total, how long did these awakenings last? [\(Show More\)](#)

2 hours 15 minutes
- What time was your final awakening? [\(Show More\)](#)

7:00 AM Fri Jan 13 2023
- What time did you get out of bed for the day? [\(Show More\)](#)

7:30 AM Fri Jan 13 2023
- How would you rate the quality of your sleep? [\(Show More\)](#)

☐ Poor
 ☐ Fair
 ☐ Good
 ☐ Very Good
- In total, how long did you nap or doze yesterday? [\(Show More\)](#)

☐ I didn't nap or doze

30 minutes
- How many drinks containing alcohol did you have? [\(Show More\)](#)

☐ No alcohol

1
- What time was your last drink? [\(Show More\)](#)

9:00 PM Thu Jan 12 2023
- Did you take any over-the-counter or prescription medication(s) to help you sleep? [\(Show More\)](#)

☐ No sleep meds

melatonin at 10:00 PM

Continue

Figure 2: SHUTi Homepage with Three Available Sleep Diaries

SHUT*i*
Sleep Healthy Understand the Mindset

Welcome to SHUTi

It's time to complete Sleep Diaries for the post assessment. Remember the diary is 10 quick questions and takes less than 5 minutes to do. When you add a diary for today's date, the questions will be about your sleep last night.

To add a diary, click the yellow Add button for an active diary date.

My Dashboard


Diaries 	You have added 0 diaries for the assessment. (Show More)	Sat Jan 7	Sun Jan 8	Mon Jan 9	Tue Jan 10	Wed Jan 11 Add >	Thu Jan 12 Add >	Fri Jan 13 Add >
---	---	------------------------	------------------------	------------------------	-------------------------	----------------------------------	----------------------------------	----------------------------------

Figure 3: Graph in SHUTi of a User's Sleep Diary Data




Figure 4: SHUTi Diary Item Queuing

SHUTi
Helping living the moment

[Help](#) [Back](#) [Info](#)

My Sleep Diary Entry For


Mon
Jun
05






Tue
Jun
06

1	What time did you get into bed? Show More	<div>Enter Time</div>
2	What time did you try to go to sleep? Show More	<input type="text"/>
3	How long did it take you to fall asleep? Show More	<input type="text"/>
4	How many times did you wake up, not counting your final awakening? Show More	<input type="text"/>


Figure 5: SHUTi “My Sleep Story” from a Completed Sleep Diary



 Help
  Back
 

My Sleep Diary Entry For

Thu
Jan
12



Fri
Jan
13

My Sleep Story

This is your Sleep Story. Please review and click the 'Edit Diary' button to make changes. Click the 'Submit Diary' button when you are done.

I got into bed at 10:00 PM on January 12, 2023

I tried to go to sleep at 10:30 PM on January 12, 2023

It took me 5 minutes to fall asleep

My final awakening was at 6:00 AM on January 13, 2023

I got out of bed for the day at 6:00 AM on January 13, 2023

I rated the quality of my sleep Fair

In total, I napped or dozed 30 minutes yesterday

I consumed 1 alcoholic drink(s) with my last drink at 9:00 PM on January 12, 2023

I took the following over-the-counter or prescription medication(s) to help me sleep melatonin at 10:00 PM

MY PERSONAL NOTES
[\(Show More\)](#)

Edit Diary

Submit Diary

Figure 6: Sleep Diary Outcomes Computation by Assessment Period from the SHUTiOASIS Trial

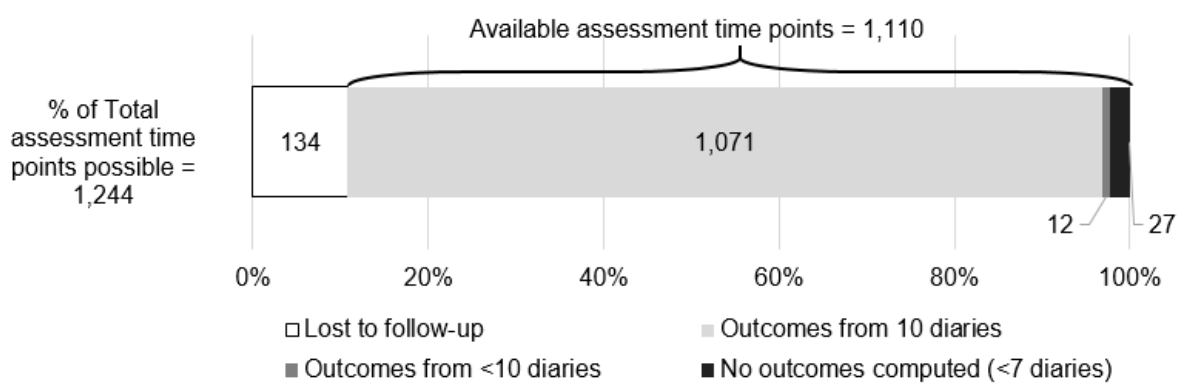


Table 1: Online Sleep Diary System Development Recommendations

Concept	Specific Recommendation	Implementation in SHUTi System
Limit retrospective bias	Restrict how late a diary may be entered	A diary may be submitted up to two days after it becomes available (e.g., a sleep diary for evening of January 1 to morning of January 2 may be entered from January 2-4, but may no longer be entered as of January 5)
Promote data entry by supporting user engagement	Provide reminders	Automated email reminders to complete diaries are sent daily
	Return value from diaries to users	Graphic data visualizations are provided
		Diary data is used to personalize and update users' sleep windows as part of sleep restriction
Promote accurate data entry through diary presentation	Complete user testing of diaries	Iterative process of interviews with experienced SHUTi diary users "Think-aloud" prototype testing of potential diary refinements with SHUTi novices
	Ensure data entry is clear and simple	Sleep diary items are queued and isolated: All questions are shown, but only one item is open for data entry at a time, while other items are "greyed out"
		Point-and-click/touch enabled interface for selecting time variables (rather than entering data freely)
		Dates associated with selected times are automatically determined by the system (although the user may override these system determinations if incorrect)
		Diary webpages are optimized for mobile devices, so may be entered on any computer, tablet, or smartphone that is Internet-connected
	Accommodate differing time-date presentation standards	Month abbreviations are used for dates rather than numerals (e.g., Jan 02 2023 instead of 01/02/2023) to limit confusion between Month-Day-Year and Day-Month-Year customs 24-hour clock time presentation and entry is available where appropriate (instead of 12-hour)
	Ensure instructions are readily available	Hyperlinks with diary instructions are available for each item
		Examples of how to enter unusual sleep situations are listed in a FAQ page
Promote data accuracy with restrictions	Incorporate data validation	Answers to all sleep metric items are required
		Bed, sleep, wake, and rise times must be consecutive to submit diary
		Total sleep time must be ≥ 0 to submit diary
		Users review their data entered for each diary on the "My Sleep Story" page and may edit data prior to submitting the diary
	Restrict changes to past diaries	Users are not personally able to edit a sleep diary after it has been submitted (administrators may edit diaries if contacted by a user)
Support data management with back-end data capture and formatting	Tag diaries with identifying information	Diaries are automatically tagged with a Diary ID number, the user ID number, the calendar date referenced by the diary, whether the diary was part of an assessment period or during the intervention period (and if so, following which Core)
	Time-stamp diary submission	Diaries are automatically time-stamped for when they were submitted by the user
	Standardize time-date formatting	Local time data for each item are formatted on the 24-hour clock (regardless of whether the user entered the data using the 12-hour or 24-hour clock through the system front-end)
		Time-date data for each item are also automatically logged in Coordinated Universal Time (UTC)

Table 2: Online Sleep Diary Data Management Recommendations

Concept	Specific Recommendation	Implementation in SHUTi Data Management Procedures
Address common and systematic time-date errors	Address date entry errors	24 hours are added to bed time when the date was presumably entered incorrectly
		24 hours are added to sleep time when the date was presumably entered incorrectly
	Address AM/PM entry errors	12 hours are subtracted from wake time when PM was presumably entered incorrectly (instead of AM)
		12 hours are subtracted from rise time when PM was presumably entered incorrectly (instead of AM)
		12 hours are added to bed time when AM was presumably entered incorrectly (instead of PM)
		12 hours are added to sleep time when AM was presumably entered incorrectly (instead of PM)
	Remove diaries with unresolvable errors following date and AM/PM changes	Diaries are removed where bed, sleep, rise, and/or wake times are no longer in chronological order
Review remaining atypical diaries	Review diaries with unusual characteristics to determine if other systematic errors exist	Diaries with long bed-to-sleep times (i.e., 8 or more hours between bed time and sleep time) are identified and reviewed
		Diaries with long total time in bed (i.e., 14 or more hours between bed time and rise time) are identified and reviewed
		Diaries with long total sleep opportunity times (i.e., 14 or more hours between sleep time and rise time) are identified and reviewed
Select diaries from the validated set for outcomes computation	Ensure diaries reflect the intended assessment period by setting an assessment window	Assessment diaries must be completed within a window of completing self-reported questionnaires (i.e., 30 days from completing baseline questionnaires; 60 days from completing follow-up time point questionnaires)
	Limit selective reporting of data by requiring a diary block	Assessment diaries must be completed within a 14-day window (inclusive)
	Summarize sleep metrics from multiple sleep diaries	Sleep metrics are averaged across a block of 10 sleep diaries – if 10 diaries are not available, metrics may be averaged from a minimum of 7 diaries

Table 3: Summary of Recommendations for Sleep Diary Data Management

Recommendation	Predicate and Presumption	Modification	Example Presumed Errant Data	Example Modified Data	Trial Diaries Affected*
Fixing date errors					
Bed time	WHEN bed time is between midnight (00:00) and 3AM (03:00) AND time in bed is 24 hours or longer; PRESUME THAT date is mis-entered	Add 24 hours to bed time	<i>Bed time = January 1 at 1:00 AM</i> Sleep time = January 2 at 1:30 AM Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM	<i>Bed time = January 2 at 1:00 AM</i> Sleep time = January 2 at 1:30 AM Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM	152 (0.003%)
Sleep time	WHEN sleep time is between midnight (00:00) and 3AM (03:00) AND total sleep opportunity time is 24 hours or longer; PRESUME THAT date is mis-entered	Add 24 hours to sleep time	<i>Bed time = January 1 at 1:00 AM</i> <i>Sleep time = January 1 at 1:30 AM</i> Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM	<i>Bed time = January 2 at 1:00 AM</i> <i>Sleep time = January 2 at 1:30 AM</i> Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM	19 (<0.001%)
Fixing AM/PM errors					
Wake time	WHEN wake time is between 3 PM (15:00) and 11:59 PM (23:59) AND time in bed is 12 hours or longer; PRESUME THAT wake time PM is mis-entered	Subtract 12 hours from wake time	Bed time = January 2 at 1:00 AM Sleep time = January 2 at 1:30 AM <i>Wake time = January 2 at 7:00 PM</i> <i>Rise time = January 2 at 7:30 PM</i>	Bed time = January 2 at 1:00 AM Sleep time = January 2 at 1:30 AM <i>Wake time = January 2 at 7:00 AM</i> <i>Rise time = January 2 at 7:30 AM</i>	36 (<0.001%)
Rise time	WHEN rise time is between 3 PM (15:00) and 11:59 PM (23:59) AND time in bed is 12 hours or longer; PRESUME THAT rise time PM is mis-entered	Subtract 12 hours from rise time	Bed time = January 2 at 1:00 AM Sleep time = January 2 at 1:30 AM Wake time = January 2 at 7:00 AM <i>Rise time = January 2 at 7:30 PM</i>	Bed time = January 2 at 1:00 AM Sleep time = January 2 at 1:30 AM Wake time = January 2 at 7:00 AM <i>Rise time = January 2 at 7:30 AM</i>	141 (0.003%)
Bed time	WHEN bed time is between 5 AM (05:00) and 3 PM (15:00) AND time in bed is 12 hours or longer; PRESUME THAT bed time PM mis-entered	Add 12 hours to bed time	<i>Bed time = January 1 at 1:00 PM</i> Sleep time = January 2 at 1:30 AM Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM	<i>Bed time = January 2 at 1:00 AM</i> Sleep time = January 2 at 1:30 AM Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM	195 (0.004%)
Sleep time	WHEN sleep time is between 5 AM (05:00) and 3 PM (15:00) AND total sleep opportunity time is 12 hours or longer; PRESUME THAT bed time PM mis-entered	Add 12 hours to bed time	<i>Bed time = January 1 at 1:00 PM</i> <i>Sleep time = January 1 at 1:30 PM</i> Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM	<i>Bed time = January 2 at 1:00 AM</i> <i>Sleep time = January 2 at 1:30 AM</i> Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM	148 (0.003%)

Table 3 cont'd

Recommendation	Predicate and Presumption	Modification	Example Presumed Errant Data	Example Modified Data	Trial Diaries Affected*
Identifying diaries with unresolvable data entry errors					
Corrected times out-of-order	[none]	[none]	Bed time = January 1 at 1:30 PM Sleep time = January 2 at 1:00 AM Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM	Bed time = January 2 at 1:30 AM Sleep time = January 2 at 1:00 AM Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM [Removed from dataset]	26 (<0.001%)
Negative total sleep (opportunity) time	[none]	[none]	Bed time = January 1 at 1:00 PM Sleep time = January 1 at 1:30 PM Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM <i>Sleep onset latency = 3 hrs</i> <i>Wake after sleep onset = 4 hrs</i> <i>Therefore total sleep time = 10.5 hrs</i>	Bed time = January 2 at 1:00 AM Sleep time = January 2 at 1:30 AM Wake time = January 2 at 7:00 AM Rise time = January 2 at 7:30 AM <i>Sleep onset latency = 3 hours</i> <i>Wake after sleep onset = 4 hrs</i> <i>Therefore total sleep time = -1.5 hrs</i> [Removed from dataset]	1 (<0.001%)
Additional diary review					
Long bed-to-sleep time	[none]	[none]	Bed time = January 1 at 11:00 PM Sleep time = January 2 at 11:00 AM Wake time = January 2 at 11:30 AM Rise time = January 2 at 11:30 AM <i>Comment: "Sick with cough"</i>	[no change]	2 (<0.001%)
Long total time in bed	[none]	[none]	Bed time = January 1 at 7:30 PM Sleep time = January 2 at 1:00 AM Wake time = January 2 at 8:30 AM Rise time = January 2 at 9:30 AM <i>Comment: "Napped late in day and never got out of bed"</i>	[no change]	31 (<0.001%)
Long total sleep opportunity time	[none]	[none]	Bed time = January 1 at 6:30 PM Sleep time = January 1 at 7:00 PM Wake time = January 2 at 10:30 AM Rise time = January 2 at 11:00 AM <i>Comment: "Extremely tired from moving son into new apartment"</i>	[no change]	18 (<0.001%)

Note: Bolded italicized lines under "Example Presumed Errant Data" and "Example Modified Data" columns denote data modified by the row-level data management recommendation. (Lines that are bolded only are presumed erroneous and modified by another data management recommendation).

*Trial information: 311 older adults with insomnia randomized to SHUTi (assessment and intervention diaries possible) or Online Patient Education (assessment diaries only) with 4 assessment time points; total diaries submitted N= 45,598 (used for percentages for date fix, AM/PM fix, and diaries with unresolvable data entry errors); total valid diaries following date and AM/PM correction and removal of diaries with unresolvable data entry errors N= 45,571 (used for percentages for additional review)

Supplemental Section: SHUTiOASIS Trial Procedures

The SHUTiOASIS trial was a fully-powered, U.S. national randomized clinical trial comparing sleep outcomes between three arms: SHUTiOASIS with a stepped-care engagement protocol, SHUTiOASIS alone, and a Patient Education control condition (primary outcomes paper in development). The sample size was determined to appropriately power testing the primary research hypotheses regarding the efficacy of the SHUTiOASIS program relative to control on key sleep outcomes. As a secondary aim, the extent to which the stepped-care engagement protocol supported program completion was examined. Sleep diaries were used both as part of trial outcomes assessment (e.g., Sleep Onset Latency and Wake After Sleep Onset were primary outcomes), as well as during the intervention period for participants randomized to either SHUTiOASIS condition.

There were four assessment time points: (1) pre-assessment, completed prior to randomization; (2) post-assessment, completed at the conclusion of the 9-week intervention period; (3) 6-month follow-up; and (4) 12-month follow-up. These assessment periods were set in order to assess sleep outcomes changes as a result of the SHUTi program (i.e., pre-post change), as well as the maintenance of those changes over time. At each assessment time point, all participants (regardless of randomized condition or completion of their assigned program) were asked to complete an online survey battery. After completion of the online surveys, participants were asked to complete 10 sleep diaries within a two-week period for the purpose of assessing outcomes. Participants received daily automated email reminders to encourage completion of these assessment sleep diaries.

Regarding intervention sleep diaries, participants randomized to SHUTi (both with and without the engagement protocol) were required to enter an additional five sleep diaries after completing Core 1 in order to “unlock” Core 2. In Core 2, sleep restriction is presented and participants receive their first personalized “sleep window,” based on their averaged total sleep time and preferred waking time. After Core 2, participants are no longer required to enter sleep diaries; however, the sleep window is only updated when additional diaries are entered.

In total, including both assessment and intervention diaries, the 311 trial participants entered more than 45,000 sleep diaries during the course of the study.