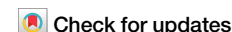




# Natural language processing in mixed-methods evaluation of a digital sleep-alcohol intervention for young adults



Frances J. Griffith<sup>1</sup> ✉, Garrett I. Ash<sup>2,3,4</sup>, Madilyn Augustine<sup>1</sup>, Leah Latimer<sup>1</sup>, Naomi Verne<sup>5</sup>, Nancy S. Redeker<sup>6</sup>, Stephanie S. O'Malley<sup>1</sup>, Kelly S. DeMartini<sup>1</sup> & Lisa M. Fucito<sup>1</sup>

We used natural language processing (NLP) in convergent mixed methods to evaluate young adults' experiences with Call it a Night (CIAN), a digital personalized feedback and coaching sleep-alcohol intervention. Young adults with heavy drinking ( $N = 120$ ) were randomized to CIAN or controls (A + SM: web-based advice + self-monitoring or A: advice; clinicaltrials.gov, 8/31/18, #NCT03658954). Most CIAN participants (72.0%) preferred coaching to control interventions. Control participants found advice more helpful than CIAN participants ( $\chi^2 = 27.34, p < 0.001$ ). Most participants were interested in sleep factors besides alcohol and appreciated increased awareness through monitoring. NLP corroborated generally positive sentiments ( $M = 15.07(10.54)$ ) and added critical insight that sleep (40%), not alcohol use (12%), was a main participant motivator. All groups had high adherence, satisfaction, and feasibility. CIAN ( $\Delta = 0.48, p = 0.008$ ) and A + SM ( $\Delta = 0.55, p < 0.001$ ) had higher reported effectiveness than A ( $F(2, 115) = 8.45, p < 0.001$ ). Digital sleep-alcohol interventions are acceptable, and improving sleep and wellness may be important motivations for young adults.

Most mental health disorders (MHDs) have a peak age of onset during or before young adulthood (18–25 years)<sup>1</sup>. However, less than half of young adults with any MHD receive treatment, with the proportion dropping to around one-third among young adults of color<sup>2</sup>. Clinical services are often unavailable or unaffordable for young adults, and the gap between treatment demand and availability has increased since the COVID-19-pandemic<sup>3</sup>. Further, young adults may not seek treatment for some highly prevalent but normalized mental health concerns, like alcohol use disorders (AUDs) and heavy (“binge”) drinking<sup>2,4,5</sup>.

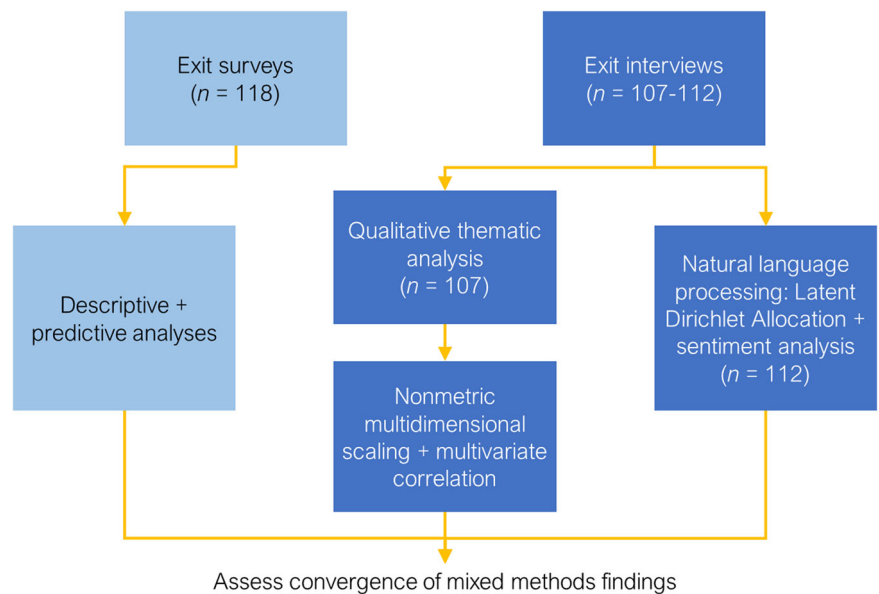
In contrast, young adults show more concern for common and related wellness issues, like sleep quality, and are more likely to seek help for their sleep concerns<sup>6,7</sup>. In a sample of young adults with similar levels of sleep complaints and heavy drinking, most (80%) were concerned about sleep, whereas very few (5%) were concerned about drinking<sup>8</sup>. Sleep and circadian disruptions, common in adolescence and young adulthood, are associated with increased alcohol use<sup>9</sup> and greater consequences from drinking<sup>10</sup>, and changes in neural reward processing are implicated in this connection<sup>11</sup>. Longitudinally, deficient sleep (i.e., sleep that is poor in quality, duration, or consistency) in adolescence, can lead to heavier drinking and AUDs later in adulthood<sup>12–14</sup>. Sleep difficulties are also linked with drinking relapse among

adults with AUDs<sup>15</sup>. Behavioral interventions for insomnia are effective for adults who report heavy drinking and show promise for reducing alcohol use or consequences<sup>14</sup>, including a digital insomnia intervention using web-based modules<sup>16</sup>. Among young adults, however, sleep difficulties are not limited to insomnia, and many struggle with poor sleep health. Therefore, a variety of sleep interventions could offer a more appealing yet effective alternative for treating young adults for sensitive issues like at-risk drinking.

Digital interventions can also increase appeal, access, and convenience of mental health treatment for young adults<sup>17</sup>. This population makes extensive use of digital health tools compared to other adults<sup>18</sup>, with generally positive experiences<sup>19</sup>. Despite a saturated market of mobile mental health apps and other digital tools, user ratings may be discrepant and uninformative to app designers and potential users alike<sup>20</sup>. Over 95% of users stop using most mobile mental health apps after 30 days<sup>21</sup>, and adherence and engagement are generally low<sup>22</sup>. In the case of sleep and alcohol apps, user engagement may increase with features like personalized feedback, interaction and support from providers, self-monitoring, and user-friendliness<sup>23–26</sup>. Further, gamification may significantly reduce attrition in mobile mental health app use in general<sup>27</sup>. With the incorporation of machine learning, apps are also becoming increasingly more personalized.

<sup>1</sup>Yale School of Medicine, Department of Psychiatry, New Haven, CT, USA. <sup>2</sup>Yale School of Medicine, Department of Biomedical Informatics and Data Science, New Haven, CT, USA. <sup>3</sup>Yale School of Medicine, General Internal Medicine, New Haven, CT, USA. <sup>4</sup>Veterans Affairs Connecticut Healthcare System, Specialty Clinics, West Haven, CT, USA. <sup>5</sup>Yale School of Public Health, Department of Social and Behavioral Science, New Haven, CT, USA. <sup>6</sup>University of Connecticut, School of Nursing, Storrs, CT, USA. ✉e-mail: [frances.griffith@yale.edu](mailto:frances.griffith@yale.edu)

**Fig. 1 | Hybrid convergent mixed methods with NLP for user experience evaluation.** Descriptive and predictive analyses were used with exit surveys. Targeted qualitative thematic analysis, post hoc NMDS and multivariate correlation, and NLP were used with exit interviews.



Assessing both users' perspectives and intervention quality is critical to promote higher treatment uptake and adherence in a rapidly growing, global digital health market<sup>28</sup>. Therefore, implementing precise, consistent evaluations of user experiences represents a significant gap in digital intervention research.

Incorporating natural language processing (NLP) into convergent mixed methods can support user-centered design for digital health interventions<sup>29</sup>. NLP is an umbrella term for a growing suite of machine learning methods that uses artificial intelligence (AI) to understand (e.g., summarize, retrieve information) and/or generate language content. Traditional evaluation methods, such as qualitative thematic analysis of exit interviews, can reveal nuance and detail, but can be impractical and unfeasible, especially with large sample sizes<sup>30</sup>. However, NLP topic modeling can help subsequent thematic analyses be more targeted and rapid. Further, exit surveys using Likert-type satisfaction scales may be subject to user response bias, and, therefore, NLP may help corroborate their results.

A hybrid approach combining NLP, quantitative survey analysis, and targeted qualitative interview analysis, may reveal broad and rich user experiences while maximizing researcher time and effort. The current study used convergent mixed methods to evaluate user experiences in a randomized clinical trial (RCT) of "Call it a Night" (CIAN) and two control conditions (see Fig. 1). CIAN is a digital personalized feedback and coaching sleep intervention for young adults that addresses heavy alcohol and other substance use. In the parent trial<sup>6</sup>, CIAN was tested against control conditions comprising either brief web-based advice only (A) or advice + active smartphone diary self-monitoring (A + SM). All participants wore sleep and alcohol biosensors for passive monitoring. NLP methods, particularly topic modeling analysis (Latent Dirichlet Allocation) and sentiment analysis, were used to assess convergence with qualitative thematic analysis of exit interviews and descriptive and predictive analysis of exit surveys. The aims of the study were to leverage these NLP evaluation methods:

1. To characterize young adults' user experiences during the digital sleep interventions for heavy drinking and
2. To determine whether user experiences varied with their demography (e.g., age, race, gender, student status, psychiatric diagnoses) or trial condition.

## Results

### Sample

We recruited 120 trial participants, and 118 completed the exit survey. Half (51%) were female and 80% were white with a mean age of 21.14 years (see

Table 1). Three-fourths (74%) were college students. Most (81%) met lifetime criteria for any MHD or substance use disorder (SUD) based on a diagnostic interview, and most also met lifetime criteria for an AUD (72%).

### Exit survey (n = 118)

On the exit survey, participants across conditions (CIAN, A + SM, and A) reported generally positive user experiences (n = 118; see Table 2). On a 5-point scale, participants reported high overall program satisfaction (M = 4.53, SD = 0.61), advice helpfulness (4.32 < M < 4.34), smartphone diary self-monitoring helpfulness (M = 4.15, SD = 0.76), and personalized feedback and coaching helpfulness (4.65 < M < 4.70). Mean feasibility ratings were generally high for the overall program (4.24 < M < 4.65), advice (M = 4.43, SD = 0.76), sleep watch biosensor (3.97 < M < 4.54), and smartphone diary self-monitoring (3.85 < M < 4.51). However, mean feasibility ratings for the ankle alcohol biosensor were lower based on visual inspection (2.41 < M < 3.91), especially due to interference with clothing, feeling uncomfortable, and being noticeable. Reported effectiveness was higher among CIAN ( $\Delta = 0.48$ ,  $p = 0.008$ ) and A + SM participants ( $\Delta = 0.55$ ,  $p < 0.001$ ) than A participants ( $F(2, 115) = 8.45$ ,  $p < 0.001$ ). Further, young adults with a lifetime history of any MHD/SUD rated their intervention as more effective on average than those without diagnoses ( $F(1, 116) = 4.64$ ,  $p < 0.001$ ;  $\Delta = 0.32$ ,  $p = 0.03$ ).

Adherence to interventions was also high across study phases. Almost all participants (98%) completed the 2-week intervention phase, and 96% completed the 12 week follow-up appointment. Regarding monitoring activities, 98% of participants wore the sleep watch biosensor for 14 days, and 95% wore the alcohol ankle biosensor for 14 days. Further, 95% of participants in the CIAN and A + SM groups completed their assigned 14 days of active smartphone diary self-monitoring of sleep and alcohol use.

### Natural language processing of exit interviews

For exit interviews, we initially used quantitative Latent Dirichlet Allocation (LDA), a topic modeling analysis that identifies the likelihood of terms occurring in topics and topics occurring in documents<sup>31</sup>, to help qualitative thematic analysis be more targeted and assess convergence of findings with exit survey analyses. Using a dataset of 112 participants' exit interviews and coaching transcripts, we identified nine topics across participants' statements (see Fig. 2; see Supplementary Materials for all topic definitions and example quotes). These topics did not vary significantly by trial condition ( $X^2 = 16.72$ ,  $p = 0.40$ ). We found that *Awareness with monitoring* was the most likely topic in the largest number of interviews (n = 18, 16%). This

**Table 1 | CIAN exit survey sample characteristics (n = 118)**

Intervention group Sample characteristic	A n (% of 29)	A + SM n (% of 29)	CIAN n (% of 60)	Total <sup>a</sup> n (% of 118)
<b>Gender</b>				
Female	14 (48.3)	16 (55.2)	30 (50.0)	60 (50.8)
Male	15 (51.7)	13 (44.8)	30 (50.0)	58 (49.2)
<b>Race</b>				
Asian	3 (10.3)	2 (6.9)	5 (8.3)	10 (8.5)
Black	1 (3.4)	1 (3.4)	7 (11.7)	9 (7.6)
White	25 (86.2)	24 (82.8)	45 (75.0)	94 (79.7)
Multiracial	0 (0.0)	0 (0.0)	1 (1.7)	1 (0.8)
Other	0 (0.0)	2 (6.9)	2 (3.3)	4 (3.4)
<b>Ethnicity</b>				
Not Hispanic, Latine	24 (82.8)	24 (82.8)	51 (85.0)	99 (83.9)
Hispanic, Latine	5 (17.2)	5 (17.2)	9 (15.0)	19 (16.1)
<b>Student status</b>				
Student	21 (72.4)	23 (79.3)	43 (71.7)	87 (73.7)
Non-student	8 (27.6)	6 (20.7)	17 (28.3)	31 (26.3)
<b>History of mental health or substance use diagnosis</b>				
Yes	23 (79.3)	25 (86.2)	48 (80.0)	96 (81.4)
No	6 (20.7)	4 (13.8)	12 (20.0)	22 (18.6)
<b>History of alcohol use diagnosis</b>				
Yes	17 (58.6)	24 (82.8)	44 (73.3)	85 (72.0)
No	12 (41.4)	5 (17.2)	16 (26.7)	33 (28.0)
AUDIT Total Score (M, SD, range) = 13.02, 4.70, 5-30 score				
Age (M, SD, range) = 21.14, 1.76, 18-25 years				

Note. "CIAN" stands for Call it a Night, the full digital sleep intervention, including personalized feedback/coaching, active smartphone diary self-monitoring, passive biosensor monitoring, and web-based advice. "A + SM" stands for advice plus self-monitoring control conditions, which also includes passive monitoring. "A" stands for advice control condition, which also includes passive monitoring. "AUDIT" stands for Alcohol Use Disorders Identification Test<sup>16</sup>. <sup>a</sup>There were no significant differences between trial groups on any demographic variable.

included participants' perceptions that both active smartphone diary self-monitoring and passive biosensor monitoring increased their awareness and mindfulness of their behaviors. As one participant stated, "Knowing that I had to do the sleep diary the next day and...people were watching what I was doing...I was just more aware of my habits."

Other common topics focused on web-based advice, especially sleep strategies. The topic, *Web advice improves sleep* (n = 16, 14%), included experiences of beginning to improve sleep using sleep strategies, such as consistent sleep schedules. A participant described, "My sleeping habits were definitely awful before now, and I think they've improved a lot, since starting this [study]...I'm starting to go to bed earlier and...reduce my activities before bed." Almost as many interviews were most likely to include the topic, *Strategy barriers* (n = 15, 13%), or descriptions of challenges implementing sleep strategies, such as situational factors (e.g., work and school schedules, dormitory environment) or personal factors (e.g., memory, motivation). The topic *Changing poor sleep* (n = 15, 13%), included descriptions of poor sleep quality or nonrestorative sleep, how this motivated study participation, and a desire to learn about different sleep factors. Therefore, three common topics focused on sleep and motivations to change sleep (a total of 40% of all interviews).

Other topics focused on participants' interest in gaining more wellness-related knowledge to change sleep or alcohol use. Thirteen interviews (12%) were most likely to include the topic *Learning and reduced drinking*, which focused on gaining new information from digital interventions and

reducing drinking given its impacts on sleep. The topic *Using feedback and web advice* (n = 11, 10%) included participants' accounts of how they integrated personalized feedback and coaching with web-based advice, including information and strategies, to alter their sleep and alcohol use. The topic *Multiple strategies and factors* (n = 10, 9%) focused on participants' curiosity and interest in the impacts of varied sleep-related factors, such as situations, environments, substances other than alcohol, and diet and exercise.

The two least frequent topics included the burden of wearable biosensors. The topic *Strategies, not devices* (n = 8, 7%) focused on helpful web-based advice aspects, attempts to implement sleep strategies, and challenges with wearable biosensors, especially the ankle alcohol biosensor. Similarly, the topic *Feedback, not devices* (n = 6, 5%) focused on the benefits of personalized feedback and coaching while also discussing difficulties with wearable biosensors.

The sentiment scores of exit interviews and transcripts (n = 112) were generally positive (M = 15.07, SD = 10.54) and ranged from -16 to 47 (see Fig. 3). Positive vs. negative sentiment scores reflect the aggregate valence of participants' word usage in exit interviews, so "0" reflects a neutral net score in an interview. Sentiment scores did not vary as a function of trial condition (CIAN, A + SM, and A; F(2, 109) = 0.78, p = 0.46). Positive sentiment scores also did not vary by demographic characteristics, including gender (F(1, 110) = 0.62, p = 0.43), age (b = 0.03, t = 0.06, p = 0.95), race (F(4, 107) = 1.23, p = 0.30), ethnicity (F(1, 110) = 0.20, p = 0.66), student status (F(3, 108) = 0.75, p = 0.53), and history of any MHD/SUD (F(1, 110) = 0.23, p = 0.63). However, the word length of each document was positively correlated with sentiment scores (r = 0.21, p = 0.03), indicating a tendency for participants who spoke at longer length to speak more positively about their experiences. To boost rigor and account for this correlation, we derived a valence-per-word score for each document and reran sentiment comparisons across condition and demographic groups with the valence-per-word score as the outcome. There were still no between-group differences (p > 0.05).

### Thematic analysis of exit interviews

Targeted qualitative thematic analysis of completed exit interviews (n = 107) was informed by the LDA. Overarching thematic categories emerged related to helpfulness of different intervention components and suggestions from participants (see Table 3 for themes and salience; see Supplementary Materials for all thematic analysis definitions and exemplar quotes).

Participants in each group condition (n = 107) highlighted various aspects of the web-based advice intervention, both helpful and unhelpful. The most mentioned helpful aspects were *General helpfulness of web-based advice* (n = 69, 64.5%) and *Helpful sleep-related information* (n = 69, 64.5%): "I think the particularly helpful ones [tips] were...things to do when you wake up, because I think...when I wake up, I just don't want to get out of bed, so I'll...stay in bed on my phone for a long time." Participants appreciated *New information from the web-based advice* (n = 58, 54.2%), *Reminders of known information* (n = 42, 39.3%), and *Usefulness of web-based advice strategies* (n = 38, 35.5%): "It [the advice] talked about...your body temperature...keeping a window open or having a fan to keep your body [cool]." Others noted helpful aspects of *Web-based advice understandability* (n = 28, 26.2%), *User-friendliness of web-based advice* (n = 25, 23.4%), *Helpful alcohol-related information* (n = 25, 23.4%), and *Memorability of information* (n = 9, 8.4%). By contrast, relatively few participants described unhelpful aspects of the advice. Some mentioned *Lack of new information* (n = 14, 13.1%) as a drawback, while others described *Irrelevance of information* (n = 5, 4.7%), such as content related to cannabis and sleep. A few participants reported *Difficulty implementing strategies* (n = 5, 4.7%) recommended through the web-based advice.

Participants across group conditions (n = 107) also provided comments on whether wearing the biosensors (ankle alcohol biosensor and sleep watch) was helpful on its own without consideration of the data collected. Most described *Neutral aspects of biosensors* (n = 65, 60.7%) and did not find them helpful or unhelpful without data. Among those who found the

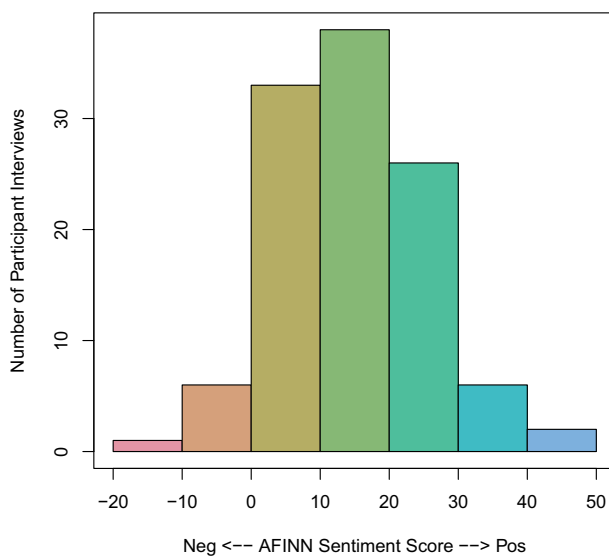
**Table 2 | CIAN exit survey outcomes (N = 118)**

Outcome (Agreement 1–5)	A M (SD)	A + SM M (SD)	CIAN M (SD)	Total M (SD)
<b>General program outcomes</b>				
Overall satisfaction	4.45 (0.74)	4.48 (0.57)	4.60 (0.56)	4.53 (0.61)
Understandability	4.52 (0.87)	4.59 (0.83)	4.65 (0.66)	4.60 (0.75)
Schedule workability	4.66 (0.94)	4.62 (0.86)	4.67 (0.82)	4.65 (0.85)
Appropriate visit length	4.31 (1.11)	4.10 (1.14)	4.27 (0.92)	4.24 (1.02)
Lifestyle change promotion	4.03 (0.87)	4.10 (0.77)	4.23 (0.96)	4.15 (0.89)
Comfortability	4.76 (0.79)	4.66 (0.81)	4.57 (0.81)	4.64 (0.80)
Hope for change promotion	4.00 (0.89)	3.97 (0.91)	4.10 (1.02)	4.04 (0.96)
Importance of target habits	4.48 (0.87)	4.52 (0.69)	4.57 (0.81)	4.53 (0.79)
Overall effectiveness	4.00 (0.76)	4.48 (0.51)	4.55 (0.57)	4.40 (0.64)
<b>Web-based advice (A) outcomes</b>				
Web-based advice module 1 helpfulness	4.24 (0.74)	4.24 (0.64)	4.43 (0.65)	4.34 (0.67)
Web-based advice module 2 helpfulness	4.07 (0.96)	4.31 (0.66)	4.45 (0.65)	4.32 (0.75)
Web-based advice user-friendliness	4.52 (0.51)	4.41 (0.50)	4.40 (0.94)	4.43 (0.76)
Web-based advice comparability to other websites	4.76 (0.44)	4.52 (0.57)	4.73 (0.58)	4.68 (0.55)
Web-based advice enjoyment	3.97 (0.91)	3.97 (0.73)	4.00 (0.77)	3.98 (0.79)
<b>Passive-monitoring (biosensor) outcomes</b>				
Sleep watch comfortable	3.93 (1.13)	4.14 (0.69)	3.92 (0.94)	3.97 (0.94)
Sleep watch not embarrassing	4.00 (1.16)	4.24 (0.64)	4.23 (0.79)	4.18 (0.86)
Sleep watch did not interfere with work	4.34 (0.97)	4.55 (0.51)	4.52 (0.62)	4.48 (0.70)
Sleep watch did not interfere with exercise	4.41 (0.78)	4.31 (0.66)	4.32 (0.91)	4.34 (0.82)
Sleep watch did not interfere with sleep	4.41 (0.68)	4.34 (0.67)	4.30 (0.89)	4.34 (0.79)
Sleep watch did not interfere with concentration	4.59 (0.57)	4.55 (0.57)	4.52 (0.60)	4.54 (0.58)
Sleep watch did not interfere with clothing	4.41 (0.87)	4.21 (0.98)	4.18 (1.02)	4.25 (0.97)
Sleep watch not noticeable	4.07 (1.00)	4.17 (0.93)	4.17 (0.96)	4.14 (0.95)
Sleep watch did not change routine	4.59 (0.57)	4.38 (0.82)	4.28 (1.01)	4.38 (0.88)
Sleep watch wear more	4.24 (0.87)	4.34 (0.81)	4.27 (0.92)	4.28 (0.88)
Ankle alcohol biosensor comfortable	2.15 (0.95)	2.66 (1.11)	2.60 (1.08)	2.51 (1.07)
Ankle alcohol biosensor not embarrassing	2.59 (1.19)	2.66 (1.42)	2.98 (1.38)	2.81 (1.35)
Ankle alcohol biosensor did not interfere with work	3.37 (1.08)	3.79 (1.05)	4.05 (1.00)	3.83 (1.06)
Ankle alcohol biosensor did not interfere with exercise	2.48 (1.16)	3.03 (1.30)	3.02 (1.32)	2.90 (1.29)
Ankle alcohol biosensor did not interfere with sleep	3.04 (1.19)	3.34 (1.14)	3.25 (1.16)	3.22 (1.16)
Ankle alcohol biosensor did not interfere with concentration	3.56 (1.25)	3.97 (0.91)	4.05 (1.02)	3.91 (1.06)
Ankle alcohol biosensor did not interfere with clothing	2.33 (1.47)	2.17 (1.28)	2.55 (1.43)	2.41 (1.40)
Ankle alcohol biosensor not noticeable	2.85 (1.22)	2.66 (1.37)	2.60 (1.24)	2.67 (1.26)
Ankle alcohol biosensor did not change routine	3.11 (1.25)	3.21 (1.18)	3.10 (1.36)	3.12 (1.28)
Ankle alcohol biosensor wear more	3.26 (1.53)	2.93 (1.25)	3.42 (1.37)	3.26 (1.38)
<b>Smartphone diary self-monitoring (SM) outcomes</b>				
Smartphone diary self-monitoring helpfulness	-	4.07 (0.72)	4.19 (0.79)	4.15 (0.76)
Smartphone diary self-monitoring easiness	-	4.34 (0.61)	4.58 (0.56)	4.51 (0.59)
Smartphone diary self-monitoring not burdensome	-	3.97 (1.09)	4.17 (0.91)	4.10 (0.97)
Smartphone diary self-monitoring did not change schedule	-	4.24 (0.83)	4.27 (1.06)	4.26 (0.98)
Smartphone diary self-monitoring easy to remember	-	3.72 (1.16)	3.92 (1.00)	3.85 (1.05)
Smartphone diary self-monitoring enjoyable	-	3.52 (0.91)	3.83 (0.91)	3.73 (0.91)
Smartphone diary self-monitoring would continue	-	3.52 (1.15)	3.83 (0.92)	3.73 (1.01)
<b>Feedback (F) from biosensor and diary monitoring outcomes</b>				
Sleep feedback helpfulness	-	-	4.70 (0.53)	4.70 (0.53)
Alcohol feedback helpfulness	-	-	4.65 (0.55)	4.65 (0.55)

Note. “CIAN” stands for Call it a Night, the full digital sleep intervention, including personalized feedback/coaching, active smartphone diary self-monitoring, passive biosensor monitoring, and web-based advice. “A + SM” stands for advice plus self-monitoring control conditions, which also includes passive monitoring. “A” stands for advice control condition, which also includes passive monitoring. Total means and standard deviations are based on 118 participants who completed the exit interview, including 60 CIAN, 29 A + SM, and 29 A participants. Some items had minor levels of missingness (<5 participants).



**Fig. 2 | LDA topic and term frequency in exit interviews.** This bar plot shows the frequency of the topics from the LDA among 112 exit interviews and feedback transcripts. The ten most common terms within each topic are shown in the order of frequency.



**Fig. 3 | Exit interview sentiment frequency.** This histogram shows the frequency of interviews’ net sentiment scores based on the AFINN lexicon<sup>46</sup> among 112 exit interviews and feedback transcripts.

biosensors helpful on their own, the most mentioned aspect was that the *Ankle alcohol biosensor increased awareness* of their alcohol use ( $n = 36, 33.6\%$ ): “The physical monitors all over you...they definitely make you... think like, “Wow...should I really be drinking this much?” Likewise, others noted the *General helpfulness of biosensors* ( $n = 35, 32.7\%$ ) or that the *Sleep watch biosensor increased awareness* ( $n = 33, 30.8\%$ ) of their sleep patterns

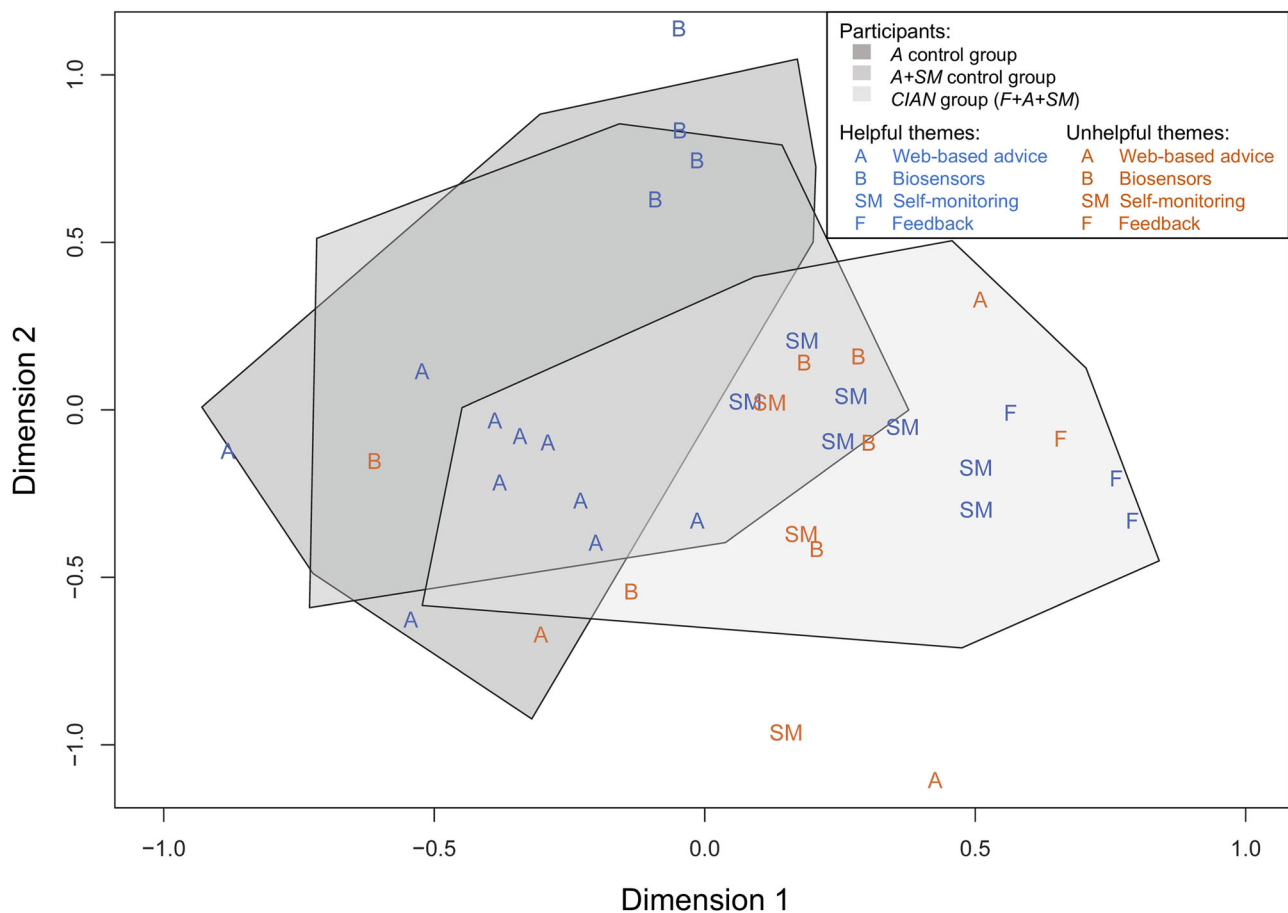
irrespective of data obtained: “If I saw my sleep watch before I was going to bed, I was like, ‘You know what, I should probably try and go to bed early tonight.’” Among those who found one or both biosensors unhelpful, the most common complaint was *Burdensomeness of ankle alcohol biosensor* ( $n = 22, 20.6\%$ ), including itchiness, discomfort, and difficulty completing some activities: “I just really didn’t like the [ankle] device...couldn’t run as much.” Other unhelpful aspects mentioned were *Unhelpful without data feedback* ( $n = 16, 15.0\%$ ), *Lack of behavior change* ( $n = 12, 11.2\%$ ), *Burdensomeness of sleep watch biosensor* (e.g., forgetting to push buttons;  $n = 8, 7.5\%$ ), *Stigma of ankle alcohol biosensor* (e.g., connection to court-ordered monitors;  $n = 6, 5.6\%$ ), and *Ankle alcohol biosensor increased drinking* ( $n = 5, 4.7\%$ ).

CIAN and A + SM participants ( $n = 78$ ) commented on the helpfulness of smartphone diary self-monitoring of sleep and alcohol use, a component of their group conditions. Most described one or more helpful aspects, most commonly that self-monitoring *Increased mindfulness of sleep* ( $n = 42, 53.8\%$ ): “I found it...helpful to...realize it [my sleep schedule] and...just reflect on...my lifestyle and...how much sleep I’m actually getting.” Others noted the *General helpfulness of self-monitoring* ( $n = 38, 48.7\%$ ) and *Increased mindfulness of alcohol use* ( $n = 31, 39.7\%$ ) “I...never really...used to like keeping track [of drinking]. So, I’d...drink a lot, and then I’d feel too drunk.” Some noted the helpfulness of the *New experience of self-monitoring* ( $n = 22, 28.2\%$ ), the *Ease of answering questions* while self-monitoring ( $n = 17, 21.8\%$ ), or the *Motivation to change behaviors* prompted by self-monitoring ( $n = 10, 12.8\%$ ). Fewer participants noted unhelpful aspects of self-monitoring, such as *Challenges answering questions* ( $n = 13, 16.7\%$ ), *Lack of behavior change* ( $n = 6, 7.7\%$ ), and *Lack of new information* (i.e., self-monitoring did not increase awareness of patterns nor motivate them to change behaviors;  $n = 6, 7.7\%$ ). Some also noted a *Neutral aspect of self-monitoring* ( $n = 13, 16.7\%$ ), they were neither helpful nor unhelpful.

**Table 3 | CIAN thematic analysis salience (N = 107)**

Thematic category	Theme	n (%) <sup>a</sup>
Helpful aspects of web-based advice	General helpfulness of web-based advice	69 ( <b>64.5%</b> of 107)
	Helpful sleep-related information	69 ( <b>64.5%</b> of 107)
	New information from the web-based advice	58 (54.2% of 107)
	Reminders of known information	42 (39.3% of 107)
	Usefulness of web-based advice strategies	38 (35.5% of 107)
	Web-based advice understandability	28 (26.2% of 107)
	Helpful alcohol-related information	25 (23.4% of 107)
	User-friendliness of web-based advice	25 (23.4% of 107)
	Memorability of information	9 (8.4% of 107)
Unhelpful aspects of web-based advice	Lack of new information	14 ( <b>13.1%</b> of 107)
	Irrelevance of information	5 (4.7% of 107)
	Difficulty implementing strategies	5 (4.7% of 107)
Helpful aspects of biosensors	Ankle alcohol biosensor increased awareness	36 ( <b>33.6%</b> of 107)
	General helpfulness of biosensor(s)	35 (32.7% of 107)
	Sleep watch biosensor increased awareness	33 (30.8% of 107)
Neutral aspect of biosensors	Neutral aspect of biosensors	65 ( <b>60.7%</b> of 107)
Unhelpful aspects of biosensors	Burdensomeness of ankle alcohol biosensor	22 ( <b>20.6%</b> of 107)
	Unhelpful without data feedback	16 (15.0% of 107)
	Lack of behavior change	12 (11.2% of 107)
	Burdensomeness of sleep watch biosensor	8 (7.5% of 107)
	Stigma of ankle alcohol biosensor	6 (5.6% of 107)
	Ankle alcohol biosensor increased drinking	5 (4.7% of 107)
Helpful aspects of smartphone diary self-monitoring	Increased mindfulness of sleep	42 ( <b>53.8%</b> of 78)
	General helpfulness of self-monitoring	38 (48.7% of 78)
	Increased mindfulness of alcohol use	31 (39.7% of 78)
	New experience of self-monitoring	22 (28.2% of 78)
	Ease of answering questions	17 (21.8% of 78)
	Motivation to change behaviors	10 (12.8% of 78)
	Neutral aspect of smartphone diary self-monitoring	Neutral aspect of self-monitoring
Unhelpful aspects of smartphone diary self-monitoring	Challenges answering questions	13 ( <b>16.7%</b> of 78)
	Lack of behavior change	6 (7.7% of 78)
	Lack of new information	6 (7.7% of 78)
Helpful aspects of feedback on biosensor and smartphone diary data	Helpfulness of personalization	45 ( <b>90.0%</b> of 50)
	General helpfulness of feedback	29 (58.0% of 50)
	Helpful data presentations	27 (54.0% of 50)
Unhelpful aspects of feedback on biosensor and smartphone diary data	Lack of new information	4 ( <b>8.0%</b> of 50)
Most influential interventions	Feedback was most influential	36 ( <b>72.0%</b> of 50)
	Web-based advice was most influential	39 (36.4% of 107)
	Self-monitoring was most influential	25 (32.1% of 78)
	Biosensors were most influential	17 (15.9% of 107)
	Multiple things were helpful	15 (14.0% of 107)
Suggestions for interventions	Willing to wear more biosensors (watch)	65 ( <b>60.7%</b> of 107)
	Want diet and exercise-related feedback	63 (58.9% of 107)
	Willing to complete new smartphone diary self-monitoring questions	60 (56.1% of 107)
	Want caffeine-related feedback	53 (49.5% of 107)
	Want other potential feedback	33 (30.8% of 107)
	Want environment-related feedback	26 (24.3% of 107)
	Unwilling to add smartphone diary question and/or biosensor	18 (16.8% of 107)
	Improvements for feedback	5 (10.0% of 50)
	Improvements for smartphone diary self-monitoring	6 (7.7% of 78)
	Improvements for biosensors	8 (7.5% of 107)
	Improvements for web-based advice	4 (3.7% of 107)

Note. Theme salience (% of those who stated the theme) is based on the number of participants who answered the interview question that resulted in the theme. General questions and questions about the web-based advice and biosensors were asked of all 107 participants across trial groups. Questions about self-monitoring were asked of the 78 CIAN and A + SM group participants who completed the self-monitoring and the exit interviews. Questions about feedback were asked of the 50 CIAN participants who received feedback and completed the exit interview. <sup>a</sup>Bold values indicate the most commonly expressed theme in a thematic category.



**Fig. 4 | NMDS ordination plot of user experience themes from exit interviews.** This ordination plot shows study themes (letters) from thematic analysis mapped on the first two dimensions of the NMDS ordination. Gray polygons were drawn around participants' locations in vector space to indicate trial groups. Themes about

helpful aspects of interventions are shown in blue, themes about unhelpful aspects of interventions are shown in orange. "A" stands for web-based advice, "B" stands for biosensors, "SM" stands for smartphone diary self-monitoring, and "F" stands for feedback/coaching.

Participants in the CIAN treatment condition ( $n = 50$ ) with personalized feedback and coaching largely described helpful aspects of the feedback and coaching intervention component. Almost all CIAN group participants described the *Helpfulness of personalization* ( $n = 45, 90.0\%$ ), the individualized nature of reports and suggestions based on biosensor and smartphone diary monitoring data: "Having that sort of feedback, both the data and just the explanation behind it, I think that's a really...unique insight into...a part of yourself that...most people wouldn't ever really get to actually see mapped out." Most also noted helpful aspects, such as the *General helpfulness of feedback* ( $n = 29, 58.0\%$ ) and *Helpful data presentations* ( $n = 27, 54.0\%$ ): "She [the coach] made it very easy for me to understand the charts and what everything meant." Only a few described one unhelpful aspect of feedback: *Lack of new information* (i.e., reports did not contribute new insights into their sleep or drinking;  $n = 4, 8.0\%$ ).

Each participant was also asked to select which received intervention component(s) were most influential for behavior change. In order from most to least helpful, participants selected personalized feedback/coaching (72.0% of the 50 CIAN participants,  $n = 36$ ), the web-based advice (36.4% of all 107 participants,  $n = 39$ ), smartphone diary self-monitoring (32.1% of the 78 CIAN and A + SM participants,  $n = 25$ ), and biosensors (15.9% of all 107 participants,  $n = 17$ ).

During exit interviews, participants also offered suggestions for current and future intervention improvement. Many participants were interested in measuring additional sleep-related factors, most commonly that they *Want diet and exercise-related feedback* ( $n = 63, 58.9\%$  of 107): "[I'd like to know] what time when you exercise, how it affects your sleep...what that does to...your whole body chemistry, everything." Others mentioned they *Want*

*caffeine-related feedback* ( $n = 53, 49.5\%$  of 107), *Want environment-related feedback* (e.g., impacts of light, noise, and temperature on sleep;  $n = 26, 24.3\%$  of 107), or *Want other potential feedback* (e.g., cognitive alertness, tobacco, and stress;  $n = 33, 30.8\%$ ): "Maybe you can see how stressed you are during... the week. Just get those numbers, and then I think that would be pretty interesting to figure [out]." To receive additional feedback, most responded that they would be *Willing to wear more biosensors* ( $n = 65, 60.7\%$ ) and *Willing to complete new smartphone diary self-monitoring questions* ( $n = 60, 56.1\%$ ). Fewer stated they were *Unwilling to add a smartphone diary question and/or biosensor* ( $n = 18, 16.8\%$ ), usually disinterested in another ankle biosensor. Several volunteered suggestions for received interventions in the current study. These included *Improvements for feedback* (e.g., combined, summary report for both intervention weeks;  $n = 5, 10.0\%$  of 50), *Improvements for smartphone diary self-monitoring* (e.g., higher pay per day;  $n = 6, 7.7\%$ ), *Improvements for biosensors* (e.g., eliminating ankle biosensors, measuring GPS location;  $n = 8, 7.5\%$ ), and *Improvements for web-based advice* (e.g., wider distribution, alternative summaries;  $n = 4, 3.7\%$ ).

### Nonmetric multidimensional scaling of interview themes

We used multivariate analysis, nonmetric multidimensional scaling (NMDS)<sup>32</sup>, to visualize the interrelationships among qualitative themes and participants. NMDS is an ordination method that condenses variation in matrices, such as participant-by-theme frequency, to a small number of orthogonal dimensions<sup>32</sup>. We used five orthogonal dimensions based on stress-level testing, and the first two dimensions are plotted in vector space in Fig. 4. Also, multivariate correlational analysis allowed us to fit participant

**Table 4 | Convergent mixed methods results**

Interventions	Exit Survey	Exit Interview		
		NLP: Topics	NLP: Sentiment	Thematic analysis
Overall program	<ul style="list-style-type: none"> <li>• <b>High satisfaction, did not vary by condition<sup>a</sup></b></li> <li>• High feasibility, did not vary by condition</li> <li>• High effectiveness (<i>CIAN</i>, <i>A + SM</i> more than <i>A</i>)</li> <li>• More effective for people with MHDs/SUDs</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Interest in holistic wellness, multiple factors</b></li> <li>• Topics did not vary by condition</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Positive sentiment, did not vary by condition or demography</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Interest in holistic wellness, multiple factors</b></li> </ul>
Feedback + coaching on biosensor and smartphone diary data ( <i>CIAN</i> group)	<ul style="list-style-type: none"> <li>• <b>Highly helpful</b></li> <li>• Highly feasible</li> </ul>	<ul style="list-style-type: none"> <li>• Integrated well with advice</li> <li>• <b>Preferred over biosensors</b></li> </ul>	-	<ul style="list-style-type: none"> <li>• <b>Most influential for three-fourths of group</b></li> <li>• <b>Helpful data personalization and feedback</b></li> <li>• Unhelpful information</li> </ul>
Smartphone diary self-monitoring ( <i>CIAN</i> , <i>A + SM</i> groups)	<ul style="list-style-type: none"> <li>• <b>Highly helpful</b></li> <li>• <b>Highly feasible</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Increased awareness</b></li> </ul>	-	<ul style="list-style-type: none"> <li>• <b>Most influential for one-third of groups</b></li> <li>• <b>Increased awareness</b></li> <li>• <b>Motivation to change</b></li> <li>• <b>Feasible</b></li> <li>• Lack of feasibility</li> <li>• Lack of change</li> <li>• Unhelpful information</li> <li>• Neutral aspects</li> </ul>
Biosensors (All groups)	<ul style="list-style-type: none"> <li>• <b>Moderately feasible (ankle alcohol biosensor)</b></li> <li>• Highly feasible (sleep watch)</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Increased awareness</b></li> </ul>	-	<ul style="list-style-type: none"> <li>• Most influential for &lt;one-fifth of all</li> <li>• <b>Increased awareness</b></li> <li>• Unhelpful data</li> <li>• <b>Burdensomeness, lack of feasibility</b></li> <li>• Lack of change</li> <li>• Neutral aspects</li> </ul>
Web-based advice (All groups)	<ul style="list-style-type: none"> <li>• <b>Highly helpful</b></li> <li>• <b>Highly feasible</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Improved sleep</b></li> <li>• <b>Challenges using strategies</b></li> <li>• Integrated well with feedback</li> <li>• <b>Preferred over biosensors</b></li> </ul>	-	<ul style="list-style-type: none"> <li>• <b>Most influential for one-third of all</b></li> <li>• <b>Helpful information and useful strategies</b></li> <li>• <b>Feasible</b></li> <li>• Unhelpful information</li> <li>• <b>Challenges using strategies</b></li> </ul>

Note. “Feedback” refers to the personalized feedback and coaching intervention components only provided as part of the Call it a Night (*CIAN*) treatment group condition. “SM” stands for smartphone diary self-monitoring, or the active use of diaries to self-report daily experiences. This intervention component was provided as part of the *CIAN* treatment group condition and the *A + SM* control condition. “Biosensors” refers to the wearable biosensor devices (sleep watch and ankle alcohol biosensor) used for passive monitoring by all trial participants across conditions. <sup>a</sup>Bold text indicates findings that converged across two or more methods.

characteristics as factors to the NMDS ordination of themes. Trial group ( $R^2 = 0.29$ ,  $p = 0.001$ ) and lifetime history of any MHD/SUD ( $R^2 = 0.03$ ,  $p = 0.04$ ) were significantly correlated with NMDS scores. That is, participants’ statements about study interventions in their exit interviews were associated with trial group and diagnostic history. Other participant characteristics (e.g., age, gender, race, ethnicity, student status) were not associated with NMDS scores. In the NMDS shown in Fig. 4, *CIAN* and *A + SM* participants were closer to themes of feedback and smartphone diary self-monitoring helpfulness than web-based advice or biosensor helpfulness in vector space, which suggests their preference for feedback and self-monitoring. Conversely, control *A* and *A + SM* participants were closer to themes of helpful web-based advice aspects in vector space than *CIAN* participants, and control participants were also significantly more likely to state these themes during interviews ( $X^2 = 27.34$ ,  $p < 0.001$ ).

### Discussion

The current study used an innovative hybrid approach in which NLP findings converged with and corroborated more traditional mixed methods to efficiently evaluate user experiences with a digital sleep intervention to reduce drinking. As shown in a joint display of mixed methods results (Table 4), findings across methods showed that participants in each condition generally found the overall program and aspects of each intervention component helpful, satisfying, and feasible. However, multiple methods showed the ankle alcohol biosensor was considered less feasible than other components, and some participants had challenges using strategies from the

web-based advice. The *CIAN* group markedly preferred the personalized feedback and coaching based on biosensor and smartphone diary data they received to other components as shown by survey effectiveness ratings and descriptions of influence in interviews. Convergent results also underscored young adults’ strong interest in gaining a more holistic picture of their wellness, including increasing their awareness through monitoring multiple health factors. Our hybrid approach using NLP enabled qualitative thematic analysis to target specific interview questions more efficiently, optimizing researcher’s time. Further, assessing the convergence of findings between NLP and other methods helped address potential researcher bias in thematic analysis and potential user response bias in exit surveys.

Mixed methods findings on user experiences, including NLP, emphasized greater effectiveness and satisfaction with personalized feedback and coaching based on biosensor and smartphone diary data than other intervention components despite generally favorable experiences with web-based advice, active smartphone diary self-monitoring, and passive biosensor monitoring. *CIAN* and *A + SM* group participants who received feedback and/or actively self-monitored rated their intervention as more effective than *A* control participants in the exit survey. Topic modeling (LDA) and thematic analysis showed that some participants found web-based advice strategies difficult to implement, and negative themes about web-based advice were associated more with *CIAN* participants.

Our study adds insight to extant literature on young adults’ preferences for interactive digital interventions, including alcohol and sleep treatment. An umbrella overview of meta-analyses and systematic reviews<sup>28</sup> found that



digital interventions offering interaction with a health professional or other social support had higher adherence and effectiveness and lower attrition. Mobile alcohol treatment apps providing personalized feedback may be especially engaging<sup>24</sup>. Further, recent research on mobile sleep apps<sup>23,25,26</sup> highlights increased engagement and satisfaction with personalized feedback and communication with health professionals. Our study adds findings that individually tailored coaching from a health professional significantly enhanced participants' experience interpreting digital health data, which likely supported their understanding. Current research also shows that user-friendliness<sup>24</sup> and evidence-based information<sup>25</sup> increase engagement, which was consistent with our finding that web-based advice was considered helpful and user-friendly, especially by control groups. Therefore, whereas web-based advice and other automated interventions are likely to be helpful to many young adults, tailored feedback and coaching with a health professional based on individualized data are likely to be preferable in the current digital health market.

Consistent with prior findings<sup>18</sup>, our results showed that young adults particularly want to use digital tools to pursue holistic health and wellness goals. Thematic analysis and LDA with interviews converged to show that participants are interested in understanding varied other lifestyle factors that impact sleep beyond alcohol (e.g., caffeine, diet, exercise, stress, environment). Most participants expressed a willingness to use additional digital tools (e.g., more biosensors, additional smartphone diary self-monitoring) to attain whole-health, personalized feedback. LDA results highlighted greater perceived awareness and mindfulness of sleep and drinking through smartphone diary self-monitoring and/or wearing biosensors. Thus, while young adults are interested in and capable of using varied digital health tools, incorporating appealing treatment foci could be important to increase young adults' generally low engagement and adherence<sup>21,22</sup>.

Our study suggests that digital interventions improving overall health may represent a more appealing treatment approach to young adults for potentially stigmatizing diagnoses, such as AUDs. Most available digital SUD interventions for young adults involve web-based advice explicitly targeting alcohol use to the exclusion of other substance use or wellness factors<sup>33</sup>. Whereas young adults may not be concerned about alcohol use or seek traditional alcohol treatment<sup>5</sup>, they are concerned about improving sleep quality<sup>6,8</sup> and other aspects of their general wellness. There have been few sleep interventions for young adults despite this connection. In this study, LDA of exit interviews showed that participants commonly joined the study to improve their sleep quality rather than address their alcohol use. Further, survey results, showed that participants with a lifetime history of an MHD/SUD found the current intervention to be especially effective. Therefore, digital interventions which focus more explicitly on improving wellness (e.g., enhancing sleep quality), like the current study, may be especially engaging to young adults with AUDs or other diagnoses who have not presented to treatment.

Our study has implications for clinical researchers and mobile app designers making user-centered app designs. User experience evaluations of digital health tools should incorporate a hybrid, convergent approach to optimize depth, breadth, and efficiency. Each user experience evaluation method has distinct strengths and potential weaknesses<sup>30</sup>. Thematic analysis of exit interviews can provide detailed information across complex subjective experiences. In the current study, along with post hoc analyses, thematic analysis revealed more distinctions between perceived helpfulness of different intervention components than any other method. However, as sample sizes increase, qualitative thematic analysis can become impractical and unfeasible. Our use of NLP topic modeling (LDA) to derive overarching topics across interview questions enabled us to target and abbreviate our thematic analysis to interview questions focusing on intervention component helpfulness. Exit surveys are also time-efficient like NLP but may be subject to user response bias. NLP, specifically sentiment analysis, helped confirm the generally positive perceptions derived from exit survey analyses.

Our convergent evaluation findings also have specific implications for digital intervention design. Young adults are enthusiastic about receiving a broad, holistic picture of factors impacting their wellness, including different

aspects of their lifestyle, in addition to appreciating individualized and specific coaching. Therefore, mobile mental health apps and linked biosensors should have a variety of choices for active and passive monitoring that can be tailored to user feasibility and interest. Digital tools should integrate seamlessly into young adults' schedules that often include exercise, which may make ankle alcohol biosensors challenging and potentially stigmatizing<sup>34</sup> compared to sleep watch biosensors. Also, digital interventions, including for sleep and alcohol use, should empower young adults to actively explore their wellness with support from a health professional. While participants in the current study valued the mindfulness derived from digital health monitoring, they preferred to also receive personalized data reports with explanations and suggested health tips from a coach. Some individuals may lack the expertise to interpret their health data or to devise action plans for improvement. Thus, digital interventions should include mechanisms for interactive feedback, such as an on-call health provider, pre-scheduled video or text check-ins with a health provider, or a chatbot that provides a spectrum of suggestions and individual data descriptions.

Digital interventions for specific, sensitive clinical issues, like heavy drinking, could engage more young adults via a focus on related general health and wellness behaviors, like sleep. Interventions that increase accessibility and engagement for young adults are important given that only half of young adults with MHDs currently receive any mental health treatment<sup>2</sup>. Digital sleep interventions for reduced drinking can target two risky, prevalent, and modifiable issues: drinking and sleep problems. Young adults commonly experience AUDs, binge drinking<sup>2</sup>, and sleep problems<sup>7</sup>. Further, sleep problems in adolescence lead to heavier drinking and AUDs in adulthood<sup>11–13</sup>. Notably, young adults who drink heavily may be less concerned about their drinking than their sleep<sup>8</sup>. In general, digital tools that explicitly focus on less sensitive behavioral goals, like improved sleep, with implications for improving potentially stigmatized goals, like reduced drinking, are likely to be more appealing and engaging to young adults.

Distinct strengths of the current study included a relatively large sample size, especially for a qualitative analysis of exit interviews; generally high adherence across intervention phases and groups; and minor amounts of missing data. Each of these attributes increased the validity of findings. Further, our hybrid, convergent approach incorporating NLP represents an important strength as these methods are promising and relatively new in digital medicine<sup>29,30</sup>. Our research also had notable limitations. Although missing data were minor, it is possible that additional themes would have emerged in a complete set of exit interviews. NLP helped address response bias in exit survey items, but survey responses were not normally distributed and generally skewed towards more positive ratings. Our sample was primarily college students, so it may not be representative of young adults in general. Further, study participants were paid, so implementation and uptake may differ in clinical or other naturalistic settings.

The current evaluation demonstrates the value of NLP in convergent mixed methods to efficiently capture broad and nuanced user experiences with digital health interventions. Specifically, our results show that digital sleep interventions for heavy drinking may increase appeal and access to alcohol treatment for young adults, especially when they include tailored coaching. Broadly, the current findings emphasize the importance of digital tools for young adults that provide a holistic, dynamic view of health coupled with interactive, individualized feedback. Consistent and precise evaluations that leverage user feedback are critical to support higher uptake and adherence to effective digital health interventions, which can begin to address the mental health treatment demand gap for young adults.

## Methods

### Study design

In the current study, we evaluated user experiences as part of an RCT of *CIAN*, a novel, two-session personalized biometric feedback and tailored coaching digital sleep intervention developed for young adults to reduce drinking ( $N = 120$ ). Participants were randomly assigned to one of three conditions using a 2:1:1 ratio for 2 weeks: (1) *CIAN*: two-session "Call it a Night" intervention ( $n = 60$ ) or (2) one of two control conditions,

A: brief advice only ( $n = 30$ ) or A + SM: brief advice + active smartphone diary self-monitoring ( $n = 30$ ). For this outpatient study in participants' natural environments, all wore sleep and alcohol biosensors daily for 2 weeks and viewed a two-module web-based sleep advice program during brief, midweek research appointments. Participants assigned to A + SM and CIAN conditions also self-monitored their sleep and alcohol use daily via smartphone diaries. Those assigned to the CIAN treatment group received personalized sleep and drinking feedback based on their biosensor and smartphone diary self-monitoring data and health coaching tailored to these data in two brief video sessions with a counselor.

There were no significant differences between the trial groups on any demographic variable (i.e., age, gender, race, ethnicity, student status, lifetime history of psychiatric diagnosis, or baseline alcohol use). Of the participants, 118 completed the exit survey and 107 completed the exit interview with study staff.

To evaluate user experiences, we used convergent mixed methods with NLP following the 2-week intervention phase. Exit surveys and interviews were conducted in parallel with participants across trial conditions to attain both broad and deep perspectives on satisfaction and intervention acceptability and feasibility. Both qualitative thematic analysis and quantitative NLP examined exit interviews. Further, in line with convergent mixed methodology, exit survey results were analyzed using descriptive and predictive analyses to compare these findings with exit interview results. This clinical trial was approved by the Institutional Review Board of Yale University (IRB# 2000021048) and was registered on clinicaltrials.gov on 8/13/2018 (NCT #03658954). The research was conducted in accordance with the Declaration of Helsinki. All participants provided informed consent. Additional study details can be found in the published study protocol paper<sup>6</sup>.

## Recruitment

To be eligible for the CIAN clinical trial, participants needed to (a) be 18–25 years of age, (b) be fluent in English, (c) self-report three or more heavy drinking occasions in the last 2 weeks (i.e., 5 or more drinks on one occasion for men or 4 or more for women), (d) score at risk of harm from drinking on the AUDIT-C<sup>35,36</sup>, and (e) self-report sleep concerns. Exclusion criteria are described further in the study protocol paper<sup>6</sup>. Participants were recruited using online (e.g., Facebook, Instagram, Snapchat) and in-person advertisements placed around the local community. Online advertisements and flyers directed individuals to a web-based pre-screening survey, and those who met pre-screening eligibility were invited to an intake visit for final eligibility determination.

## Exit survey and interview

Upon completion of the 2-week intervention phase, we asked participants to complete a self-report exit survey. The exit survey included Likert-type and yes/no questions to assess user experiences, including overall satisfaction, intervention helpfulness and effectiveness, and intervention feasibility. Participants were also asked to detail areas for improvement in open-ended response items.

Participants across trial groups also completed semi-structured exit interviews to provide information about their subjective user experiences (see Supplementary Materials for the exit interview protocol and exit survey items). Exit interviews were administered by three research staff members (two men, one woman), who were trained in the protocol and familiar with study interventions. The exit interview protocol included questions about the following topics: participants' sleep and alcohol use before and after the study, perceptions of these behaviors relative to their peers, use of sleep hygiene strategies from the intervention, helpfulness of intervention components, the impact of payment on study participation, what drew them to the study, and suggestions for intervention improvement or future studies targeting sleep or alcohol use.

## Data analysis

We conducted both qualitative thematic analysis and quantitative NLP with exit interviews. For thematic analysis, we used Braun and Clarke's six-step process<sup>37</sup>, during which we used a recursive and iterative team coding process to ensure rigor and trustworthiness in the process of identifying, naming, and categorizing recurrent themes. Two members of the research team completed an initial open coding of the entire text of a large subset of exit interviews ( $n = 80$ ) to derive initial themes. Then, guided by findings from NLP topic modeling analysis, the current thematic analysis targeted key portions of the exit interviews, including questions related to helpfulness of intervention components and participants' suggestions. Three research team members engaged in a rigorous team coding process of the entire dataset of completed 2-week exit interviews ( $n = 107$ ). Using an initial codebook based on previous first-round, open thematic coding, these three researchers independently coded a random selection of 20% of the entire interview dataset ( $n = 21$ ) to evaluate interrater reliability. All themes had percent agreement of 95% or greater, but an extensive reconciliation meeting was still undertaken to audit the initial codebook and resolve any disagreements until an overall kappa exceeding 0.7 was reached between each pair of coders. The remaining exit interviews were divided among the three researchers to code using the revised codebook. Auditing was used throughout the qualitative process to ensure transparency, including field notes maintained by each coder and reviewed by other coders and ongoing check-ins with the entire research team.

To visualize the interrelationship between participants and themes, we conducted NMDS with theme frequency counts using the *vegan* package<sup>31</sup> in the programming language R<sup>38</sup>. NMDS is a multivariate ordination method that condenses variation in matrices to a small number of orthogonal dimensions<sup>32</sup>. We selected NMDS for post-hoc analysis of themes because it does not place assumptions of normality on frequency count data compared to other multidimensional scaling methods<sup>32</sup>. To assess the goodness of fit when selecting the number of orthogonal dimensions, we used stress level cutoffs and a stress plot (Shepard diagram)<sup>31</sup>. We selected five orthogonal dimensions, resulting in a good to fair stress score of .13, which is likely to result in minimal distortion<sup>39</sup>. As shown in Fig. 4, we used a multivariate correlation analysis with *vegan* in R to fit factors (trial condition and participant gender, race, ethnicity, student status, and diagnosis history) and a vector (participant age) to the NMDS ordination of participants and themes. This enabled us to assess whether participants' user experiences (as interview themes) varied significantly with their assigned trial condition or personal characteristics.

For NLP of exit interviews, we conducted topic modeling<sup>40</sup> and sentiment analyses<sup>41</sup> with an expanded dataset of all of participants' statements during completed exit interviews and feedback session transcripts ( $n = 112$ ). LDA with the *topicmodels* package<sup>40</sup> in R was used for topic modeling analysis. To boost rigor and ensure the most parsimonious number of topics ( $k$ ) were used for LDA, we undertook a preliminary analysis with three model selection methods<sup>42–44</sup>, to test different  $k$  values in the *ldatuning* package<sup>45</sup> in R and ultimately selected nine topics. Text in participants' statements was preprocessed before NLP, including removing punctuation, numbers, capitalization, and English 'stopwords' as well as stemming all remaining words. The LDA determined which of the nine topics was most likely to occur in each document and which terms were most likely to occur in each topic. Topics were systematically named by closely reading five or more interview transcripts determined most likely to contain the topic by LDA. Chi-squared tests determined whether topics varied with participants' assigned trial condition.

For sentiment analysis, we used the AFINN lexicon<sup>46</sup> and the *syuzhet* package<sup>41</sup> in R. The AFINN lexicon is a well-established and commonly used sentiment lexicon<sup>46</sup> that assigns a valence score from -5 to 5 to selected terms, and documents are scored based on an aggregate of their terms' scores. An aggregate AFINN sentiment score of "0" for a document indicates neutral term valence. Bivariate regression models, including analysis of variance (ANOVA), were used to assess whether participants' trial

condition or personal characteristics predicted the sentiment of their statements during exit interviews.

Descriptive and predictive analyses were conducted with exit survey results. These included summary statistics of each item and linear modeling to determine if participants' trial condition or personal characteristics predicted responses to exit survey items.

### Data availability

The datasets used and analyzed during the current study are available from the corresponding author and the principal investigator, Dr. Fucito, on reasonable request.

### Code availability

The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Received: 21 February 2024; Accepted: 30 October 2024;

Published online: 29 November 2024

### References

- Solmi, M. et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol. Psychiatr.* **27**, 281–295 (2022).
- SAMHSA & Center for Behavioral Health Statistics and Quality. *Results From The 2022 National Survey on Drug Use and Health*. <https://www.samhsa.gov/data/report/2022-nsduh-detailed-tables> (2023)
- Moreno, C. et al. How mental health care should change as a consequence of the COVID-19 pandemic. *Lancet Psychiatr* **7**, 813–824 (2020).
- Lee, M. R. & Sher, K. J. “Maturing out” of binge and problem drinking. *Alcohol Res.* **39**, 31–42 (2018).
- Singleton, R. A. Jr. & Wolfson, A. R. Alcohol consumption, sleep, and academic performance among college students. *J. Stud. Alcohol Drugs* **70**, 355–363 (2009).
- Fucito, L. M. et al. A multimodal mobile sleep intervention for young adults engaged in risky drinking: protocol for a randomized controlled trial. *JMIR Res. Protoc.* **10**, e26557 (2021).
- Sivertsen, B. et al. Sleep patterns and insomnia in young adults: a national survey of Norwegian university students. *J. Sleep. Res.* **28**, e12790 (2019).
- Ash, G. I. et al. Using web-based social media to recruit heavy-drinking young adults for sleep intervention: prospective observational study. *JMIR* **22**, e17449 (2020).
- Hasler, B. P., Schulz, C. T. & Pedersen, S. L. Sleep-related predictors of risk for alcohol use and related problems in adolescents and young adults. *Alcohol Res* **44**, 02 (2024).
- Miller, M. B., DiBello, A. M., Lust, S. A., Carey, M. P. & Carey, K. B. Adequate sleep moderates the prospective association between alcohol use and consequences. *Addict. Behav.* **63**, 23–28 (2016).
- Hasler, B. P., Soehner, A. M. & Clark, D. B. Sleep and circadian contributions to adolescent alcohol use disorder. *Alcohol* **49**, 377–387 (2015).
- Hasler, B. P., Martin, C. S., Wood, D. S., Rosario, B. & Clark, D. B. A longitudinal study of insomnia and other sleep complaints in adolescents with and without alcohol use disorders. *Alcohol Clin. Exp. Res.* **38**, 2225–2233 (2014).
- Hasler, B. P., Kirisci, L. & Clark, D. B. Restless sleep and variable sleep timing during late childhood accelerate the onset of alcohol and other drug involvement. *J. Stud. Alcohol Drugs* **77**, 649–655 (2016).
- Mike, T. B., Shaw, D. S., Forbes, E. E., Sitnick, S. L. & Hasler, B. P. The hazards of bad sleep—Sleep duration and quality as predictors of adolescent alcohol and cannabis use. *Drug Alcohol Depend* **168**, 335–339 (2016).
- Brooks, A. T. & Wallen, G. R. Sleep disturbances in individuals with alcohol-related disorders: a review of cognitive-behavioral therapy for insomnia (CBT-I) and associated non-pharmacological therapies. *Subst. Abuse.* **8**, 55–62 (2014).
- Verlinden, J. J. et al. Effects of a digital cognitive behavioral therapy for insomnia on sleep and alcohol consumption in heavy drinkers: a randomized pilot study. *Alcohol Clin. Exp. Res.* **47**, 2354–2365 (2023).
- Domhardt, M. et al. Mechanisms of change in digital health interventions for mental disorders in youth: systematic review. *JMIR* **23**, e29742 (2021).
- Paradis, S., Roussel, J., Bosson, J. L. & Kern, J. B. Use of smartphone health apps among patients aged 18 to 69 years in primary care: population-based cross-sectional survey. *JMIR Formative Res.* **6**, e34882 (2022).
- Clement, A., Ravet, M., Stanger, C. & Gabrielli, J. Feasibility, usability, and acceptability of MobileCoach-Teen: a smartphone app-based preventative intervention for risky adolescent drinking behavior. *J. Subst. Use Addiction Treat.* **159**, 209275 (2024).
- Carlo, A. D., Hosseini Ghomi, R., Renn, B. N. & Areán, P. A. By the numbers: ratings and utilization of behavioral health mobile applications. *Npj Digital Med.* **2**, 1–8 (2019).
- Baumel, A., Muench, F., Edan, S. & Kane, J. M. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *JMIR* **21**, e14567 (2019).
- Garrido, S. et al. What works and what doesn't work? a systematic review of digital mental health interventions for depression and anxiety in young people. *Front Psychiatr* **10**, 759 (2019).
- Aji, M. et al. Exploring user needs and preferences for mobile apps for sleep disturbance: Mixed methods study. *JMIR Ment. Health* **6**, e13895 (2019).
- Crane, D., Garnett, C., Brown, J., West, R. & Michie, S. Factors influencing usability of a smartphone app to reduce excessive alcohol consumption: think aloud and interview studies. *Front Public Health* **5**, 240357 (2017).
- Lee-Tobin, P. A., Ogeil, R. P., Savic, M. & Lubman, D. I. Rate my sleep: examining the information, function, and basis in empirical evidence within sleep applications for mobile devices. *J. Clin. Sleep. Med.* **13**, 1349–1354 (2017).
- Oyebode O., Alhasani M., Mulchandani D., Olagunju T., Orji R. SleepFit: A persuasive mobile app for improving sleep habits in young adults. In *2021 IEEE 9th Int. Conference on Serious Games and Applications for Health (SeGAH)*. 1–8 (IEEE, 2021).
- Litvin, S., Saunders, R., Maier, M. A. & Lüttke, S. Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: a randomized controlled trial. *PLoS ONE* **15**, e0237220 (2020).
- Lehtimäki, S., Martic, J., Wahl, B., Foster, K. T. & Schwalbe, N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. *JMIR Ment. Health* **8**, e25847 (2021).
- Petersen, C. L. et al. Using natural language processing and sentiment analysis to augment traditional user-centered design: development and usability study. *JMIR Mhealth Uhealth* **8**, e16862 (2020).
- Skeen, S. J., Jones, S. S., Cruse, C. M. & Horvath, K. J. Integrating natural language processing and interpretive thematic analyses to gain human-centered design insights on HIV mobile health: proof-of-concept analysis. *JMIR Hum. Factors* **9**, e37350 (2022).
- Oksanen J. et al. *Vegan: Community Ecology Package. R Package Version 2.6-4*. <https://CRAN.R-project.org/package=vegan> (2022)
- Ding C. S. *Fundamentals of applied multidimensional scaling for educational and psychological research* 1st ed. 2018 edition, Vol. 197 (Springer International Publishing, 2018).
- Monarque, M., Sabetti, J. & Ferrari, M. Digital interventions for substance use disorders in young people: rapid review. *Subst. Abuse. Treat. Prev. Policy* **18**, 13 (2023).

34. Ash, G. I. et al. specificity, and tolerability of the BACTrack skyn compared to other alcohol monitoring approaches among young adults in a field-based setting. *Alcohol Clin. Exp. Res.* **46**, 783–796 (2022).
35. Demartini, K. S. & Carey, K. B. Correlates of AUDIT risk status for male and female college students. *J. Am. Coll. Health* **58**, 233–239 (2009).
36. Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R. & Grant, M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption II. *Addiction* **88**, 791–804 (1993).
37. Braun, V. & Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **3**, 77–101 (2006).
38. R. Core Development Team. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.* <https://www.R-project.org/> (2004)
39. Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**, 117–143 (1993).
40. Grün B., Hornik K. *Topicmodels: Topic Models. R package version 0.2-14.* <https://CRAN.R-project.org/package=topicmodels> (2023)
41. Jockers M. L. *Syuzhet: Extract Sentiment and Plot Arcs From Text.* <https://github.com/mjockers/syuzhet> (2015)
42. Arun R., Suresh V., Veni Madhavan C. E., Narasimha Murthy M. N. On finding the natural number of topics with latent dirichlet allocation: some observations. In, *Advances in Knowledge Discovery and Data Mining. PAKDD 2010.* (eds. Zaki, M. J., Yu, J. X., Ravindran, B., Pudi, V.) 6118. (Berlin, Heidelberg, 2010).
43. Griffiths T. L., Steyvers M., Blei D. M., Tenenbaum J. B. Integrating topics and syntax. *Adv. Neural Inform. Process. Syst.* **17**, 13–18 (2004).
44. Cao, J., Xia, T., Li, J., Zhang, Y. & Tang, S. A density-based method for adaptive LDA model selection. *Neurocomputing* **72**, 1775–1781 (2009).
45. Nikita M. *ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters.* <https://CRAN.R-project.org/package=ldatuning> (2020)
46. Nielsen, F. A. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In *Proc. ESWC2011 Workshop on 'Making Sense of Microposts': Big Things Come in Small Packages 718 in CEUR Workshop Proceedings.* 93–98 (IEEE, 2011).

## Acknowledgements

This research was supported by the National Institute of Health and the National Institute on Alcohol Abuse and Alcoholism under Grants # R34AA026021, R34AA026021-S1, and R01AA03013 and the National Institute on Drug Abuse under Grant # T32DA019426. Further, we would like to thank Daniel Griffith for his consultation on statistical analyses conducted in R.

## Author contributions

F.J.G.: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review and editing. G.I.A.: Conceptualization, Methodology, Investigation, Project administration, Writing—review and editing. M.A.: Data curation, Formal analysis, Validation, Writing—review and editing. L.L.: Formal analysis, Validation, Writing—review and editing. N.V.: Formal analysis. N.S.R.: Conceptualization, Methodology, Writing—review and editing. S.S.O.: Conceptualization, Methodology, Writing—review and editing. K.S.D.: Conceptualization, Methodology, Writing—review and editing. L.M.F.: Conceptualization, Funding acquisition, Methodology, Investigation, Resources, Supervision, Validation, Writing—review and editing

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41746-024-01321-3>.

**Correspondence** and requests for materials should be addressed to Frances J. Griffith.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024