Linguistic and non-linguistic markers of disorganization in psychotic illness

Einat Liebenthal[1,2], **Michaela Ennis**[1,3], **Habiballah Rahimi-Eichi**[1,2], **Eric Lin**[1,4], **Yoonho Chung**[1], **Justin T. Baker**[1,2]

[1]McLean Institute for Technology in Psychiatry, McLean Hospital, Belmont, MA, USA

[2]Department of Psychiatry, Harvard Medical School, Boston, MA, USA

[3]Division of Medical Sciences, Harvard Medical School, Boston, MA, USA

[4]Medical Informatics, Veterans Affairs Boston, Boston, MA, USA

Abstract

Background: Disorganization, presenting as impairment in thought, language and goal-directed behavior, is a core multidimensional syndrome of psychotic disorders. This study examined whether scalable computational measures of spoken language, and smartphone usage pattern, could serve as digital biomarkers of clinical disorganization symptoms.

Methods: We examined in a longitudinal cohort of adults with a psychotic disorder, the associations between clinical measures of disorganization and computational measures of 1) spoken language derived from monthly, semi-structured, recorded clinical interviews; and 2) smartphone usage pattern derived via passive sensing technologies over the month prior to the interview. The language features included speech quantity, rate, fluency, and semantic regularity. The smartphone features included data missingness and phone usage during sleep time. The clinical measures consisted of the Positive and Negative Symptom Scale (PANSS) conceptual disorganization, difficulty in abstract thinking, and poor attention, items. Mixed linear regression analyses were used to estimate both fixed and random effects.

Results: Greater severity of clinical symptoms of conceptual disorganization was associated with greater verbosity and more disfluent speech. Greater severity of conceptual disorganization was also associated with greater missingness of smartphone data, and greater smartphone usage during

---

Corresponding Author: Einat Liebenthal, McLean Imaging Center, McLean Hospital, 115 Mills Street, Belmont, MA 02478, USA, eliebenthal@mclean.harvard.edu, office phone: 617-855-2939.

sleep time. While the observed associations were significant across the group, there was also significant variation between individuals.

Conclusions: The findings suggest that digital measures of speech disfluency may serve as scalable markers of conceptual disorganization. The findings warrant further investigation into the use of recorded interviews and passive sensing technologies to assist in the characterization and tracking of psychotic illness.

Introduction

Disorganization, presenting as an impairment in thought, language and communication (conceptual disorganization) and non-goal directed behavior (behavioral disorganization), is one of the core syndromes of psychotic disorders, including schizophrenia and bipolar disorder (Morgan et al., 2017; Yalincetin et al., 2017). Conceptual disorganization comprises difficulties in the coherent sequencing of thoughts, which can manifest as increases in typical features of spoken language such as verbosity, and atypical features such as illogical, derailed or tangential speech, distractible speech, and peculiar use of words and sentence constructions (Andreasen, 1986; Kuperberg, 2010; Liddle et al., 2002). Conceptual disorganization may also be associated with impairments in other domains of cognitive and executive functioning, including attention, memory, and abstract thinking, although the precise nature of these complex associations is not well understood and remains an area of active investigation (Bora et al., 2019; Brune and Bodenstein, 2005; Fusar-Poli et al., 2012; Klingberg et al., 2006; Vignapiano et al., 2019; Wallwork et al., 2012). In addition to disorganized speech (a positive symptom), psychotic illness may involve impoverished speech (a negative symptom), presenting as a reduction in the rate and quantity of words, sentences, and content of speech. In schizophrenia, it has been suggested that the disorganization and impoverished dimensions of spoken language co-occur in early stages of the disease, but disorganization diminishes while impoverishment persists with the progression of the disease (Palaniyappan, 2021; Roche et al., 2016).

The multidimensional and dynamic nature of language disturbances has stressed the need for objective markers, and was to a large extent the driving force behind the emergence of computational language methods, for characterizing psychotic illness (Corcoran et al., 2020; Girard et al., 2021). With respect to conceptual disorganization, most of the computational work in spoken language has focused on the difficult problems of quantifying semantic incoherence and irregularity, and syntactic complexity (Bearden et al., 2011; Bedi et al., 2015; Corcoran et al., 2018; Elvevag et al., 2007; Murphy and Ongur, 2022; Silva et al., 2022). However, disturbances in verbosity and verbal fluency, such as atypical use of pauses, verbal and non-verbal fillers, and word repeats or stutters, have also proven useful for characterizing psychotic illness (Cokal et al., 2019; de Boer et al., 2020; Tang et al., 2021). While significant advances in natural language processing (NLP) methods have facilitated the discovery of interesting observations on disrupted language in psychosis, there remain significant challenges particularly with respect to quantifying spoken language. Some contributing factors are that most NLP tools were developed based on written and not spoken language. In addition, transcription of spoken language is often not done verbatim, such that instances of disfluency may not be recorded or analyzed. Finally, many studies use

brief speech production tasks (e.g., describing a picture or comic strip), which may limit the generalizability of their findings to natural conversational speech.

Behaviors outside the domains of language and cognition are not traditionally considered in the assessment of behavioral disorganization. Nevertheless, it is conceivable that a lack of structure and routine in cyclic (daily or otherwise) activities may also reflect or contribute to behavioral disorganization. In particular, abnormal sleep patterns, including increased sleep latency and decreased sleep time and sleep efficiency, are an important characteristic of psychotic disorders (Chan et al., 2017; Chouinard et al., 2004; Zanini et al., 2013). Sleep disturbances are common in psychosis and have been associated with increased severity of clinically-relevant symptoms and neurocognitive deficits (Cohrs, 2008; Davies et al., 2017; Poe et al., 2017), although a specific association with conceptual disorganization has to our knowledge not been reported. The importance of sleep abnormalities in psychotic (and other mental) disorders, and short-comings of self-reported measures of sleep (Lauderdale et al., 2008; Silva et al., 2007), have fueled an interest in using smartphones to passively monitor sleep (Staples et al., 2017). Emerging work suggests that objective measures of sleep may be inferred from sensor data collected passively via a smartphone application, including accelerometry and screen on/off logs (Staples et al., 2017). While passive smartphone-based measures can only provide indirect, coarse estimates of sleep, the high prevalence of smartphone ownership, widespread usage of smartphones in multiple aspects of routine functioning, and unobtrusive and objective nature of passive smartphone sensing, make this technology promising for assessing sleep, particularly in longitudinal and large-scale studies.

The overall objective of the present study was to examine whether scalable computational measures of spoken language and smartphone usage pattern could serve as digital biomarkers of the latent construct of conceptual disorganization. To this end, we tested in a longitudinal cohort of adults with a psychotic disorder, the associations between clinical measures of disorganization and computational measures of 1) spoken language derived from monthly, semi-structured, recorded clinical interviews; and 2) smartphone usage pattern derived passively over the month prior to the interview. For spoken language, in order to probe the various types of disturbances that have previously been associated with disorganized and non-goal directed thought patterns (Corcoran et al., 2020; deBoer et al., 2020; Kuperberg et al., 2010), we analyzed features that index the quantity, rate, and fluency of speech, as well as features that index semantic coherence and regularity. The measures of speech verbosity and fluency, in addition to being automatically and precisely quantifiable from verbatim transcription, may be particularly informative for the assessment of conversational speech in relatively long clinical interviews as analyzed here. For the smartphone, we analyzed usage during the sleep period as a rough indicator of the quality of sleep, because sleep disturbances have consistently been associated with multiple symptoms of psychotic illness (Chouinard et al., 2004; Cohrs, 2008). We also analyzed phone data missingness as an indicator of noncompliance with the study requirements (see Methods), to examine the possibility of an association with psychotic illness symptoms. We used the Positive and Negative Symptom Scale (PANSS(Kay et al., 1987)) conceptual disorganization (P2) item (scored based on information obtained during the interviews) as the primary clinical measure of disorganization, because this assessment is intended to capture disrupted thought processes expressed in spoken language. In addition, prior factor analyses of the

PANSS have suggested that P2 has the highest loading in a disorganization factor in multiple independent studies (Wallwork et al., 2012). For thoroughness, the analyses were also conducted with the PANSS N5 (abstract thinking) and G11 (poor attention) items. These items, although relatively unspecific to the core features of disorganization in psychosis, have also consistently been assigned to a disorganization factor in multiple independent studies, albeit with lower loadings (Wallwork et al., 2012).

Methods

Study procedures

Study participants were adults who have been diagnosed with either a primary psychotic disorder (i.e., schizophrenia, schizoaffective disorder, or psychotic disorder not otherwise specified) or a psychotic condition secondary to an affective disorder (e.g., bipolar or major depressive disorder with psychotic features). Participants learned about the study via advertisement in clinical programs of the divisions for Psychotic Disorders and Depression and Anxiety Disorders at McLean Hospital, and the Rally with Mass General Brigham (MGB) platform. The study was designed to last one year in each participant, with optional extension dependent on clinical status, and a few of the participants were active for up to five years. Participation required installation of the Beiwe smartphone application (Onnela and Rauch, 2016; Torous et al., 2016) for semi-continuous, passive collection of phone usage, accelerometry, and location data. Participants could optionally complete daily microsurveys and audio journals, deployed via the Beiwe application. There was also the option of participating in monthly recorded clinical interviews, approximately 30 minutes long, during which participants were asked about their symptoms in the past month. The interviews were conducted either onsite (in the years 2016-2020) and recorded as described in our prior work (Girard et al., 2021), or virtually (since March 2020) using the MGB mandated Zoom video conferencing and recording platform. Completion of study procedures was monitored daily by study staff using an in-house developed deep phenotyping dashboard (DPdash), and feedback to participants regarding missing data, or assistance solving technical issues, were provided within 1-3 days. Participants were paid monthly according to the study procedures they completed. The protocol was approved by the MGB Institutional Review Board. The present report focuses on the associations between spoken language features and clinical symptom scores extracted from the recorded interviews, as well as the associations between smartphone use features summarized over the month prior to the interview and the clinical symptom scores derived from the interviews.

Clinical symptom scores

We focused here on the clinical symptoms of disorganization, as assessed using the PANSS (Kay et al., 1987) conceptual disorganization (P2), difficulty in abstract thinking (N5), and poor attention (G11) items, because these items have been associated with a disorganized/ concrete factor in a meta-analysis of multiple prior factor analyses (Wallwork et al., 2012). The clinical interviews conducted in this study were semi-structured and designed to obtain the information needed to score the PANSS, and a few other scales not considered here, based on questions and brief conversation with participants about symptoms and events incurred in the month prior to the interview. Each interview was scored on all PANSS items

by a trained clinical rater, with a subset of the interviews reviewed periodically by a second trained rater to detect and correct any inconsistencies.

Spoken language features

In this report, we analyzed only the participants' side of the interview. We focused on spoken language features that can be derived automatically from an interview or conversation, and have previously been suggested to be impaired in association with disorganized and non-goal directed thought patterns, including measures of speech quantity, rate, and fluency, and measures of semantic coherence and regularity (Corcoran et al., 2020; de Boer et al., 2020; Kuperberg, 2010). The audio recordings of the clinical interviews were automatically checked for quality using the Opensmile software, and recordings flagged for low sound amplitude or high interviewer-interviewee channel correlation were reviewed by study staff to determine adequacy for transcription. Audio recordings were sent to a professional service (TranscribeMe, Inc.) to obtain research grade transcriptions for analysis. The transcription was verbatim and included speaker annotation, sentence level time stamps, and redaction of personally identifying information. In addition, filler and crutch words, and word repetitions, were offset with commas, and false starts and stutters were marked systematically with dashes. We derived the spoken language features automatically from the transcripts using Python 3.9 pandas, numpy, and nltk packages, including counts of word and disfluency types, and measures of semantic processing. Semantic coherence was computed using the Google News 300 dimension word2vec model (Mikolov et al., 2013), based on the mean cosine similarity between word embeddings, for both sequential and pairwise words in a sentence. Word uncommonness was computed using the same package, based on the magnitude of each word vector. All the features (except total words and word rate) were computed at the sentence level and then averaged for each speaker over the interview. Because the transcriptions were verbatim and time-stamped at the sentence level, they accurately captured spoken sentence units, rather than strictly formal sentences as defined in written language. From a practical standpoint, sentences were the most accessible linguistic unit available for analysis. Averaging by sentence effectively adjusted the features for the different length of each participant's interview responses. The extracted features were reviewed by a staff member for accuracy in a subset of the interviews. The list and description of the analyzed language features is provided in Table 1.

Phone features

We focused here on phone usage features that could be indicative of behavioral disorganization and lack of structure in daily routines, including phone data missingness and phone usage during sleep time. Passive collection of phone usage data via the Beiwe application was a required study procedure. Phone data missingness was considered to reflect non-compliance with study requirements when it occurred in isolation (i.e., not in relation to a system error observed across participants) and could not be resolved promptly with the assistance of study staff. Common reasons for phone data missingness in an individual participant were failure to charge the phone battery, failure to connect the phone to Wi-Fi, failure to update the phone settings to permit data transfer, and uninstalling the Beiwe Application. The phone data missingness and phone in-use during sleep time features were computed using our in-house developed deep phenotyping pipeline (DPSleep). Phone

data were operationally defined as missing on any day with less than 60 minutes of phone in-use time recorded in Beiwe. The phone usage data missingness feature was computed as the number of days in which phone data were missing in the 28 days preceding the interview (or the period since the previous interview if less than 28 days). The participants' phone usage behavior was calculated only for interviews for which phone usage data missingness did not exceed 25% of the days preceding the interview. The phone usage behavior was computed based on the phone's "locked" and "unlocked" events recorded in Beiwe. The phone in-use time was computed as the time between consecutive unlocked-locked events, and was operationally limited to a maximum duration of 15 minutes (reflecting the assumption that the phone was not used continually during the unlocked-locked period). To determine sleep time, phone usage time was first computed minute by minute for every 24-hour cycle, and the primary sleep episode, operationally defined as the longest epoch in a cycle when the phone is locked, was computed using a 150-minutes moving window (Rahimi-Eichi et al., 2021). The average sleep epoch over the entire course of study participation was then computed. The number of minutes of phone in-use during the average sleep epoch was computed daily, and the mean and standard deviation over the pre-interview search period were derived for analysis in this report. The list and description of analyzed smartphone features is provided in Table 1.

Statistical Analysis

The analyses were designed to estimate the associations between each language feature and each disorganization clinical measure, and between each phone feature and each disorganization clinical measure. Because the features and clinical scores were derived from a varying number of interviews per participant, linear mixed effects models were used to estimate both fixed and random effects. Each clinical measure was regressed on each feature in a separate mixed effects linear regression model, taking participant as a random variable (allowing for different intercepts and slopes for each participant). To minimize the risk of type I errors, a Bonferroni correction was applied to account for the number of language features (n=11) and phone features (n=4) that were tested with each clinical measure. Thus, the statistical significance threshold was set at $p<0.0045$ for the models of language features, and $p<0.0125$ for the models of phone features. To assess the possibility of shared variance between language features, and between phone features, we also ran multivariate linear mixed effects models with P2 as the dependent variable and two predictors representing the primary language or phone, respectively, features found to be significantly associated with P2 in the separate mixed effects models.

Results

Seventy-four participants were enrolled in the study in total since February 2016. The demographic distribution of each sub sample used in the analyses is reported in Table 2. The full sample used for the analyses of associations with P2 scores consisted of data linked to 745 clinical interviews, obtained from 59 participants who agreed to the recorded interviews, in the period till March 2022. Several of the interviews did not have N5 or G11 scores, resulting in 737 and 743 interviews in the full samples for these clinical scores, respectively. The full samples were used to evaluate phone usage data missingness. For the analysis of

the phone usage features, the samples consisted of data linked to 471 interviews (with less than 25% missingness of phone usage data) for P2, and 470 for N5 and G11, from 51 participants. For the analysis of language features, the samples consisted of 145 interviews for P2, and 144 interviews for N5 and G11, from 18 participants. Because of the cost of the professional transcription service, transcriptions were initially obtained only for this subset of interviews, to first evaluate its potential utility. The number of available interviews varied between participants. For the full sample, the number of available interviews with P2 scores per participant ranged 1-58 (Mean, M=12.42, Standard Deviation, SD=12.46), for the phone usage sample it ranged 1-37 (M=9.24, SD=8.16), and for the language sample it ranged 1-20 (M=8.06, SD=6.92).

Clinical Scores

The distribution of clinical scores in each of the samples is shown in Figure 1. Overall, symptoms ranged from absent (score=1) to moderate severe (score=5), with one instance of severe (score=6). In the full sample, the P2 scores ranged 1-5 (M=1.43, SD=0.89), for N5 they ranged 1-6 (M=1.41, SD=0.80), and for G11 they ranged 1-5 (M=1.70, SD=0.99). Of the 59 participants included in the full sample, 25 had P2 scores 3, 26 had N5 scores 3, and 39 had G11 scores 3, in at least one interview. In the phone usage sample, the scores for P2 scores ranged 1-4 (M=1.25, SD=0.65), for N5 they ranged 1-5 (M=1.33, SD=0.72), and for G11 they ranged 1-5 (M=1.62, SD=0.92). Of the 51 participants in the phone sample, 22 had P2 scores 3, 24 had N5 scores 3, and 34 had G11 scores 3, in at least one interview. In the language sample, the P2 scores ranged 1-5 (M=2.01, SD=1.36), for N5 they ranged 1-6 (M=1.7, SD=1.03), and for G11 they ranged 1-5 (M=1.96, SD=1.22). Of the 18 participants in the language sample, 11 had P2 scores 3, 11 had N5 scores 3, and 15 had G11 scores 3, in at least one interview.

Associations of spoken language features and clinical scores

The duration of the interviews analyzed for language features ranged 2.9 to 95 minutes (M=30.01, SD=16.96). The results of the mixed linear regression analyses between each of the eleven language features and the P2 scores are shown in Figure 2. Six of the features were found to have a significant (Bonferroni corrected) positive association with P2 scores, including 'ratio of participant words' (Beta, B=0.26, 95% confidence interval, CI=±0.16, p=0.002), 'words per sentence' (B=0.27, CI=±0.17, p=0.003), 'total disfluencies per sentence' (B=0.28, CI=±0.13, p=0.0001), 'verbal edits per sentence' (B=0.20, CI=±0.13, p=0.003), 'repeats per sentence' (B=0.26, CI=±0.16, p=0.0001), and 'restarts per sentence' (B=0.26, CI=±0.12, p=0.00003). The association of 'words per second' with P2 did not survive Bonferroni correction (B=0.18, CI=±0.15, p=0.02). The other features were not significantly associated with P2 (p>0.19). There were also significant random effects of participant on the associations of all the language features with the P2 scores, suggesting that even though the fixed effects reported above were significant across the group, there was also considerable variation between participants.

Figure 3 shows scatter plots and linear fit lines for the relationships between the 'ratio of participant words' and P2, and 'total disfluencies per sentence' and P2. These two features represent the two domains of language -- verbosity and disfluency -- found to

be significantly associated with P2 in the mixed linear regression analyses with single predictors. A multivariate linear mixed effects model with P2 as the dependent variable and these two features as predictors revealed a significant effect of 'total disfluencies per sentence' (B=0.24, CI=±0.19, p=0.01), but not of 'ratio of participant words' (B=0.06, CI=±0.22, p=0.56), suggesting that there was shared variance between the two features.

None of the language features were significantly associated with the N5 (p>0.14), or G11 (p>0.15), scores.

Associations of phone features and clinical scores

The results of the mixed linear regression analyses between each of the four phone features and the P2 scores are shown in Figure 4. Two of the features were found to have a significant positive association with P2 scores, including 'phone missing days' (B=0.11, CI=±0.06, p=0.0002), and 'mean phone in-use during the sleep period' (B=0.13, CI=±0.10, p=0.01). The other features were not significantly associated with P2 (p>0.1). There were also significant random effects of participant on the associations of all the phone features with the P2 scores, suggesting again that the reported fixed effects were significant across the group, but there was also considerable variation between participants.

Figure 5 shows scatter plots and linear fit lines for the relationships between 'phone missing days' and P2, and between 'mean phone in-use during the sleep period' and P2. A multivariate linear mixed effects model with P2 as the dependent variable and these two features as predictors revealed significant effects of both (phone missing days: B=0.11, CI=±0.065, p=0.0009; mean phone in-use during the sleep period: B=0.13, CI=±0.1, p=0.01), suggesting an independent association of each feature with P2.

None of the features were significantly associated with the N5 scores (p>0.08), or the G11scores (p>0.04.

Discussion

We analyzed transcripts of semi-structured clinical interviews, and smartphone usage data, in a longitudinal cohort of adults with a psychotic disorder, to identify potential associations between computational measures derived from these data sources and clinical symptoms of disorganization. Our mixed linear regression analyses revealed specific positive associations between measures of speech quantity and disfluency, and the clinical scores representing the severity of conceptual disorganization. There were also positive associations between measures of smartphone data missingness, and smartphone usage during sleep time, and the clinical scores of conceptual disorganization. While these associations were significant across the group, there was also significant variation between individuals in the group. Overall, the results highlight the tremendous potential of basic computational linguistic and smartphone measures for supplementing the clinical evaluation of conceptual disorganization. The results also highlight the potential value of longitudinal study, as it may permit assessing how well the observed associations generalize across individuals, and may eventually aid in the disentangling of disease subtypes and stage of disease progression (Palaniyappan, 2021; Roche et al., 2016). Below, we discuss the results in more details.

Spoken language features

Across individuals, greater severity of conceptual disorganization symptoms was associated with greater verbosity (more words in total over the entire interview, more words per sentence), and more disfluent speech (more verbal edits, repeats and stutters, and restarts, per sentence, see Table 1 for definitions and examples of the features). Greater verbosity and greater speech disfluency are thought to reflect difficulties in the planning and formulation of sentences, related to the difficulties in coherent sequencing of thoughts characteristic of conceptual disorganization (Andreasen, 1986; Liddle et al., 2002). Overall, the findings are in line with other nascent work demonstrating the potential utility of quantitative analysis of speech verbosity and disfluency for characterizing psychotic illness (Cokal et al., 2019; de Boer et al., 2020; Girard et al., 2021; Tang et al., 2021). Our multivariate analysis further showed that verbosity covaried with disfluency and did not contribute additional predictive value for the association with conceptual disorganization, suggesting that the count of speech disfluencies provided an independent estimate of conceptual disorganization.

In contrast, the clinical scores of conceptual disorganization were not associated with any of the semantic regularity measures, which are intended to directly capture sentence incoherence (including instances of illogical, derailed or tangential speech, and peculiar use of words and sentence constructions). Several factors may have contributed to this null finding. First, the range of conceptual disorganization in this participant sample extended to moderate severe (defined as frequent irrelevances, disconnectedness, or loosening of associations), but not to severe (defined as seriously derailed, internally inconsistent, gross, almost constant irrelevancies) or extreme (termed as "word salad"). Because our measures consisted of summaries across the entire interview, frequent but not constant instances of incoherent sentences may have resulted in a relatively diluted incoherence score. Second, our analyses focused on sentence level features which could be computed automatically and scaled to interviews with different structures and lengths. Thus, we did not measure between-sentences or narrative level incoherence. In addition, because the interviewer's side of the interview was not analyzed, instances of participant responses that were incoherent specifically with respect to the interviewer's utterance would have been missed. Finally, a general drawback of many algorithms for computing semantic coherence and irregularity, including the word2vec model used here, is that they are trained on a corpus of written text, and may not be adjusted to handle certain sentence constructions or word short forms (e.g., 'wanna', 'cause') that occur more commonly in casual spoken language. Thus, more work is needed to further optimize and automatize semantic regularity features for the analysis of semi-structured dyadic interviews. Overall, these results are consistent with those of Tang and colleagues (Tang et al., 2021), who found in open-ended interviews of individuals with schizophrenia and absent-to-mild language disorder versus healthy controls, a large discrepancy in speech disfluency (specifically, rate of word incompletion), and a discrepancy in sentence incoherence only when participant responses were analyzed with respect to interviewer's prompts.

While across the group there were significant positive associations between the verbosity and disfluency features and the clinical measures of conceptual disorganization, there was also significant variation in these associations between individuals in the group. One source

of individual variability, specifically the different number of interviews per participant, was inherent to the study design. However, other sources of individual variability such as different ranges of conceptual disorganization scores may have been related to different severity or stage of disease. First, it is notable that the measures of verbosity and disfluency varied in association with clinically observable changes in conceptual disorganization within individuals, suggesting that these linguistic features may function as markers of varying symptom severity rather than markers of a stable disease phenotype. These results reinforce those of recent work showing that a linguistic measure of semantic similarity varied not only between individuals with psychosis and healthy controls, but also within the psychosis group in relation to observable changes in clinical symptoms (Alonso-Sanchez et al., 2022). Furthermore, the present findings of individual differences in the patterns of association between language features and clinical measures of conceptual disorganization highlight the potential value of longitudinal study of spoken language disturbances for identifying illness subtypes and tracking illness progression. For example, dominance of impoverished (negative) versus disorganized (positive) speech disturbances, which is hypothesized to occur at more advanced disease stages (Chouinard et al., 2004; Zanini et al., 2013), would manifest as a negative association with speech verbosity and disfluency counts (rather than a positive association as observed here on average across the group).

Overall, the results raise the possibility that counts of speech disfluencies constitute simple, readily automatable, markers of conceptual disorganization, perhaps especially in individuals with moderate-severe symptoms who are naturally more challenging to evaluate. The verbatim speech transcriptions obtained in this study, which included systematic marking of disfluencies by type, trivialized the automatic derivation of disfluency features, albeit at a cost that prohibited transcription of the entire sample. With the advent of improved automated speech transcription platforms, obtaining such features may further be streamlined and become more easily scalable.

Phone usage features

Across individuals, greater severity of conceptual disorganization symptoms was associated with greater missingness of smartphone usage data, and greater use of smartphone during sleep time (see Table 1 for definitions of the features). The multivariate analysis showed that each of these features contributed independent predictive value to the association with conceptual disorganization. As detailed in the Methods, missingness of phone usage data could be due to a variety of reasons, all broadly related to smartphone usage that is inconsistent with study instructions, and therefore potentially indicative of behavioral disorganization and inability to maintain routine tasks. The present finding of a positive association between conceptual disorganization and phone data missingness is consistent with the possibility that phone usage patterns could serve as a useful biomarker of behavioral routines. Patterns of smartphone usage have also been shown to provide useful information regarding daily routine behaviors such as sleep (Staples et al., 2017). In this context, increased phone use during the average sleep period could be taken to reflect reduced sleep quality. The present finding of a positive association between smartphone use during sleep time and conceptual disorganization is novel, though generally consistent with the notion of abnormal sleep patterns in psychotic disorders (Chouinard et al., 2004;

Zanini et al., 2013). While the relationship between reduced sleep quality and conceptual disorganization has not been examined directly, several prior studies have found that sleep deprivation is associated with reduced speech fluency and alterations in speech production (Harrison and Horne, 1997; Vogel et al., 2010), consistent with the possibility that disrupted sleep contributes to conceptual disorganization. Future work using standardized measurements of sleep should formally evaluate the relationships of sleep patterns with conceptual disorganization.

Despite their admitted coarseness, estimation of behavioral routines and sleep patterns from smartphone usage confer the advantage of being completely passive and placing little-to-no burden on the participant. Thus, smartphone-based passive sensing is highly applicable to probing sleep and other routine behaviors longitudinally and naturalistically. The present findings, while completely novel and awaiting replication and validation, highlight the tremendous potential value of passive sensing of behavior for identifying markers of complex syndromes such as disorganization.

Conclusions

We found that longitudinal study of individuals with a psychotic disorder, using recorded clinical interviews and passive sensing smartphone-based methods, yielded useful scientific insights. The study's main findings, of positive associations between clinical measures of conceptual disorganization and 1) measures of spoken language disfluency, and 2) measures of smartphone usage during sleep, have important implications. The findings suggest that digital measures of speech disfluency may serve as scalable markers of conceptual disorganization. The findings warrant further investigation into the use of recorded interviews and passive sensing technologies to assist in the characterization and tracking of psychotic illness. The study also has several limitations. The mixed effects linear regression analyses estimated random participant effects. However, demographic (age, sex, education) and clinical (diagnosis, medication) variables could have important effects on the observed associations and should be examined in future studies with larger participant samples. In addition, the participant sample ranged in clinical symptoms from absent to moderate severe, such that it is unclear whether the results would generalize to individuals with severe symptoms. The spoken language and smartphone usage features analyzed here were found to be specifically associated with clinical measures of conceptual disorganization, but not other facets of disfunction that have been linked to disorganization, namely, abstract thinking and poor attention. Future longitudinal studies with larger participant samples and using thought disorder specific scales should further examine the relationships between these symptom dimensions, in the context of individual variability in their manifestation over the course of disease progress.

Acknowledgements

Role of Funding

References

Alonso-Sanchez MF, Ford SD, MacKinley M, Silva A, Limongi R, Palaniyappan L, 2022. Progressive changes in descriptive discourse in First Episode Schizophrenia: a longitudinal computational semantics study. Schizophrenia (Heidelb) 8(1), 36. [PubMed: 35853894]

Andreasen NC, 1986. Scale for the assessment of thought, language, and communication (TLC). Schizophr Bull 12(3), 473–482. [PubMed: 3764363]

Bearden CE, Wu KN, Caplan R, Cannon TD, 2011. Thought disorder and communication deviance as predictors of outcome in youth at clinical high risk for psychosis. J Am Acad Child Adolesc Psychiatry 50(7), 669–680. [PubMed: 21703494]

Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, Ribeiro S, Javitt DC, Copelli M, Corcoran CM, 2015. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr 1, 15030. [PubMed: 27336038]

Bora E, Yalincetin B, Akdede BB, Alptekin K, 2019. Neurocognitive and linguistic correlates of positive and negative formal thought disorder: A meta-analysis. Schizophr Res 209, 2–11. [PubMed: 31153670]

Brune M, Bodenstein L, 2005. Proverb comprehension reconsidered--'theory of mind' and the pragmatic use of language in schizophrenia. Schizophr Res 75(2-3), 233–239. [PubMed: 15885515]

Chan MS, Chung KF, Yung KP, Yeung WF, 2017. Sleep in schizophrenia: A systematic review and meta-analysis of polysomnographic findings in case-control studies. Sleep Med Rev 32, 69–84. [PubMed: 27061476]

Chouinard S, Poulin J, Stip E, Godbout R, 2004. Sleep in untreated patients with schizophrenia: a meta-analysis. Schizophr Bull 30(4), 957–967. [PubMed: 15954201]

Cohrs S, 2008. Sleep disturbances in patients with schizophrenia : impact and effect of antipsychotics. CNS Drugs 22(11), 939–962. [PubMed: 18840034]

Cokal D, Zimmerer V, Turkington D, Ferrier N, Varley R, Watson S, Hinzen W, 2019. Disturbing the rhythm of thought: Speech pausing patterns in schizophrenia, with and without formal thought disorder. PLoS One 14(5), e0217404. [PubMed: 31150442]

Corcoran CM, Carrillo F, Fernandez-Slezak D, Bedi G, Klim C, Javitt DC, Bearden CE, Cecchi GA, 2018. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry 17(1), 67–75. [PubMed: 29352548]

Corcoran CM, Mittal VA, Bearden CE, R EG, Hitczenko K, Bilgrami Z, Savic A, Cecchi GA, Wolff P, 2020. Language as a biomarker for psychosis: A natural language processing approach. Schizophr Res 226, 158–166. [PubMed: 32499162]

Davies G, Haddock G, Yung AR, Mulligan LD, Kyle SD, 2017. A systematic review of the nature and correlates of sleep disturbance in early psychosis. Sleep Med Rev 31, 25–38. [PubMed: 26920092]

de Boer JN, Voppel AE, Brederoo SG, Wijnen FNK, Sommer IEC, 2020. Language disturbances in schizophrenia: the relation with antipsychotic medication. NPJ Schizophr 6(1), 24 [PubMed: 32895389]

Elvevag B, Foltz PW, Weinberger DR, Goldberg TE, 2007. Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. Schizophr Res 93(1-3), 304–316. [PubMed: 17433866]

Fusar-Poli P, Deste G, Smieskova R, Barlati S, Yung AR, Howes O, Stieglitz RD, Vita A, McGuire P, Borgwardt S, 2012. Cognitive functioning in prodromal psychosis: a meta-analysis. Arch Gen Psychiatry 69(6), 562–571. [PubMed: 22664547]

Girard JM, Vail AK, Liebenthal E, Brown K, Kilciksiz CM, Pennant L, Liebson E, Ongur D, Morency LP, Baker JT, 2021. Computational analysis of spoken language in acute psychosis and mania. Schizophr Res.

Harrison Y, Horne JA, 1997. Sleep deprivation affects speech. Sleep 20(10), 871–877. [PubMed: 9415947]

Kay SR, Fiszbein A, Opfer LA, 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophrenia bulletin 13(2), 261. [PubMed: 3616518]

Klingberg S, Wittorf A, Wiedemann G, 2006. Disorganization and cognitive impairment in schizophrenia: independent symptom dimensions? Eur Arch Psychiatry Clin Neurosci 256(8), 532–540. [PubMed: 17164977]

Kuperberg GR, 2010. Language in schizophrenia part 1: an introduction. Language and linguistics compass 4(8), 576–589. [PubMed: 20936080]

Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ, 2008. Self-reported and measured sleep duration: how similar are they? Epidemiology 19(6), 838–845. [PubMed: 18854708]

Liddle PF, Ngan ET, Caissie SL, Anderson CM, Bates AT, Quested DJ, White R, Weg R, 2002. Thought and Language Index: an instrument for assessing thought and language in schizophrenia. Br J Psychiatry 181, 326–330. [PubMed: 12356660]

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J, 2013. Distributed representations of words and phrases and their compositionality. Advances in Neural Information Processing Systems.

Morgan CJ, Coleman MJ, Ulgen A, Boling L, Cole JO, Johnson FV, Lerbinger J, Bodkin JA, Holzman PS, Levy DL, 2017. Thought Disorder in Schizophrenia and Bipolar Disorder Probands, Their Relatives, and Nonpsychiatric Controls. Schizophr Bull 43(3), 523–535. [PubMed: 28338967]

Murphy M, Ongur D, 2022. Thought disorder is correlated with atypical spoken binomial orderings. NPJ Schizophr 8(1), 25.

Onnela JP, Rauch SL, 2016. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. Neuropsychopharmacology 41(7), 1691–1696. [PubMed: 26818126]

Palaniyappan L, 2021. Dissecting the neurobiology of linguistic disorganisation and impoverishment in schizophrenia. Semin Cell Dev Biol.

Poe SL, Brucato G, Bruno N, Arndt LY, Ben-David S, Gill KE, Colibazzi T, Kantrowitz JT, Corcoran CM, Girgis RR, 2017. Sleep disturbances in individuals at clinical high risk for psychosis. Psychiatry Res 249, 240–243. [PubMed: 28126579]

Rahimi-Eichi H, Coombs Iii G, Vidal Bustamante CM, Onnela JP, Baker JT, Buckner RL, 2021. Open-source Longitudinal Sleep Analysis From Accelerometer Data (DPSleep): Algorithm Development and Validation. JMIR Mhealth Uhealth 9(10), e29849. [PubMed: 34612831]

Roche E, Lyne J, O'Donoghue B, Segurado R, Behan C, Renwick L, Fanning F, Madigan K, Clarke M, 2016. The prognostic value of formal thought disorder following first episode psychosis. Schizophr Res 178(1-3), 29–34. [PubMed: 27639419]

Silva AM, Limongi R, MacKinley M, Ford SD, Alonso-Sanchez MF, Palaniyappan L, 2022. Syntactic complexity of spoken language in the diagnosis of schizophrenia: A probabilistic Bayes network model. Schizophr Res.

Silva GE, Goodwin JL, Sherrill DL, Arnold JL, Bootzin RR, Smith T, Walsleben JA, Baldwin CM, Quan SF, 2007. Relationship between reported and measured sleep times: the sleep heart health study (SHHS). J Clin Sleep Med 3(6), 622–630. [PubMed: 17993045]

Staples P, Torous J, Barnett I, Carlson K, Sandoval L, Keshavan M, Onnela JP, 2017. A comparison of passive and active estimates of sleep in a cohort with schizophrenia. NPJ Schizophr 3(1), 37. [PubMed: 29038553]

Tang SX, Kriz R, Cho S, Park SJ, Harowitz J, Gur RE, Bhati MT, Wolf DH, Sedoc J, Liberman MY, 2021. Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. NPJ Schizophr 7(1), 25. [PubMed: 33990615]

Torous J, Kiang MV, Lorme J, Onnela JP, 2016. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Ment Health 3(2), e16. [PubMed: 27150677]

Vignapiano A, Koenig T, Mucci A, Giordano GM, Amodio A, Altamura M, Bellomo A, Brugnoli R, Corrivetti G, Di Lorenzo G, Girardi P, Monteleone P, Niolu C, Galderisi S, Maj M, Italian Network for Research on, P., 2019. Disorganization and cognitive impairment in schizophrenia:

New insights from electrophysiological findings. Int J Psychophysiol 145, 99–108. [PubMed: 30898578]

Vogel AP, Fletcher J, Maruff P, 2010. Acoustic analysis of the effects of sustained wakefulness on speech. J Acoust Soc Am 128(6), 3747–3756. [PubMed: 21218906]

Wallwork RS, Fortgang R, Hashimoto R, Weinberger DR, Dickinson D, 2012. Searching for a consensus five-factor model of the Positive and Negative Syndrome Scale for schizophrenia. Schizophr Res 137(1-3), 246–250. [PubMed: 22356801]

Yalincetin B, Bora E, Binbay T, Ulas H, Akdede BB, Alptekin K, 2017. Formal thought disorder in schizophrenia and bipolar disorder: A systematic review and meta-analysis. Schizophr Res 185, 2–8. [PubMed: 28017494]

Zanini M, Castro J, Coelho FM, Bittencourt L, Bressan RA, Tufik S, Brietzke E, 2013. Do sleep abnormalities and misaligned sleep/circadian rhythm patterns represent early clinical characteristics for developing psychosis in high risk populations? Neurosci Biobehav Rev 37(10 Pt 2), 2631–2637. [PubMed: 24096189]
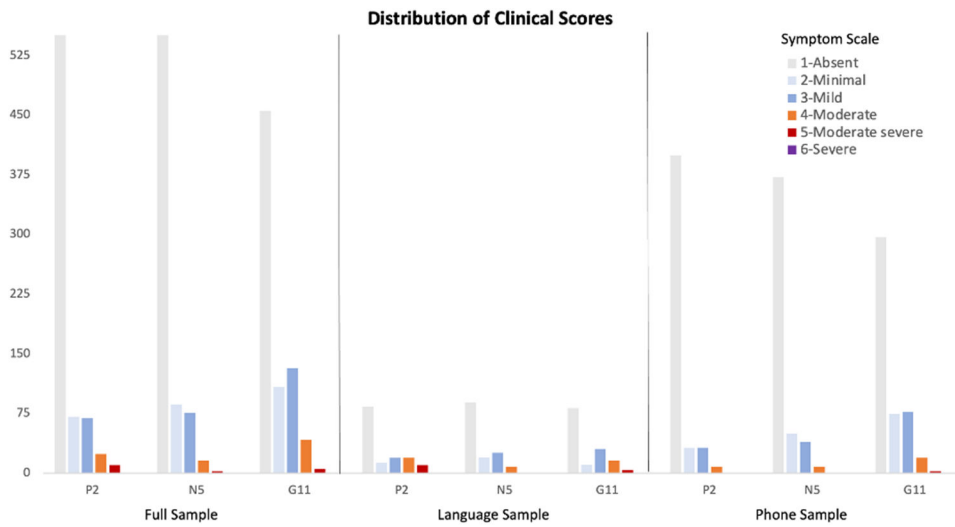
Figure 1.
Distribution of the PANSS P2 (conceptual disorganization), N5 (difficulty in abstract thinking), and G11 (poor attention) clinical scores in the full sample of interviews, and the samples used for the analysis of language, and phone, features. The color scale represents the severity of symptoms in each PANSS item. See text for definitions of the three samples.

Language features and conceptual disorganization

Figure 2.
Forest plot of beta coefficients and 95% confidence intervals for the general linear regression models representing the relationship between each language feature and the conceptual disorganization clinical score (PANSS P2). Each feature was tested in a separate mixed effects model, including participant as a random variable. Plot points colored in red mark significant (Bonferroni corrected), and plot points colored in black mark non-significant, models.

A. Ratio of Participant Words by P2 Conceptual Disorganization
Y=0.462+0.0353*X
(n=145)

B. Disfluencies per Sentence by P2 Conceptual Disorganization
Y=0.7123+0.1357*X
(n=145)

Figure 3.
A. Scatter plot of the ratio of 'total words produced by the participant over total words produced in the interview' versus the PANSS P2 conceptual disorganization scores, overlaid with the linear fit line and 95% confidence interval for the correlation between these variables. **B.** Scatter plot of the 'mean number of disfluencies per sentence' versus the PANSS P2 conceptual disorganization scores, overlaid with the linear fit line and 95% confidence interval for the correlation between these variables.
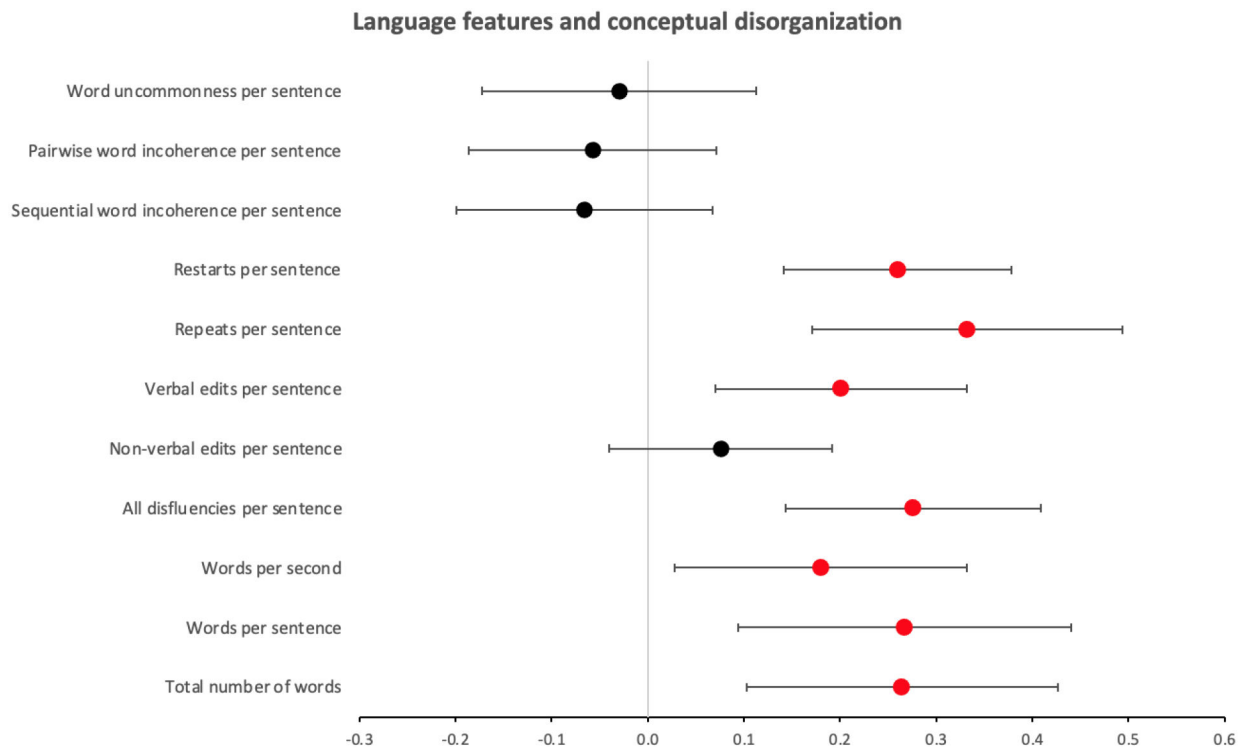
Figure 4.
Forest plot of the beta coefficients and 95% confidence intervals for the general linear regression models representing the relationship between each phone feature and the conceptual disorganization clinical scores (PANSS P2). Each feature was tested in a separate mixed effects model, including participant as a random variable. Plot points colored in red mark significant (Bonferroni corrected), and plot points colored in black mark non-significant, models.
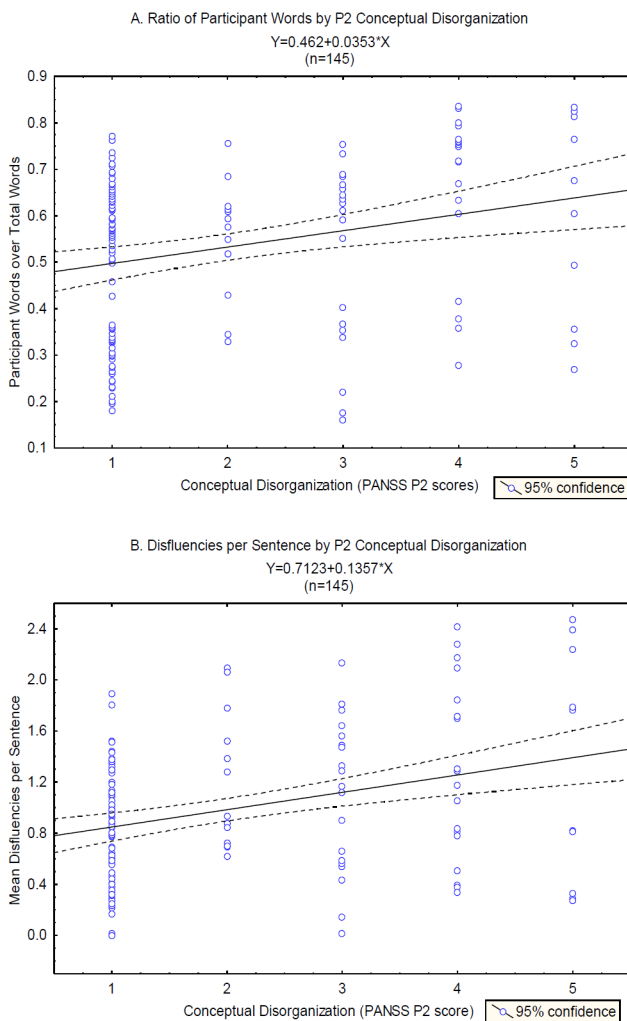
A. Phone Data Missingness by P2 Conceptual Disorganization
$Y=4.3131+2.42*X$
$(n=745)$

B. Phone In-Use During Sleep Epoch by P2 Conceptual Disorganization
$Y=9.6292+5.0425*X$
$(n=471)$
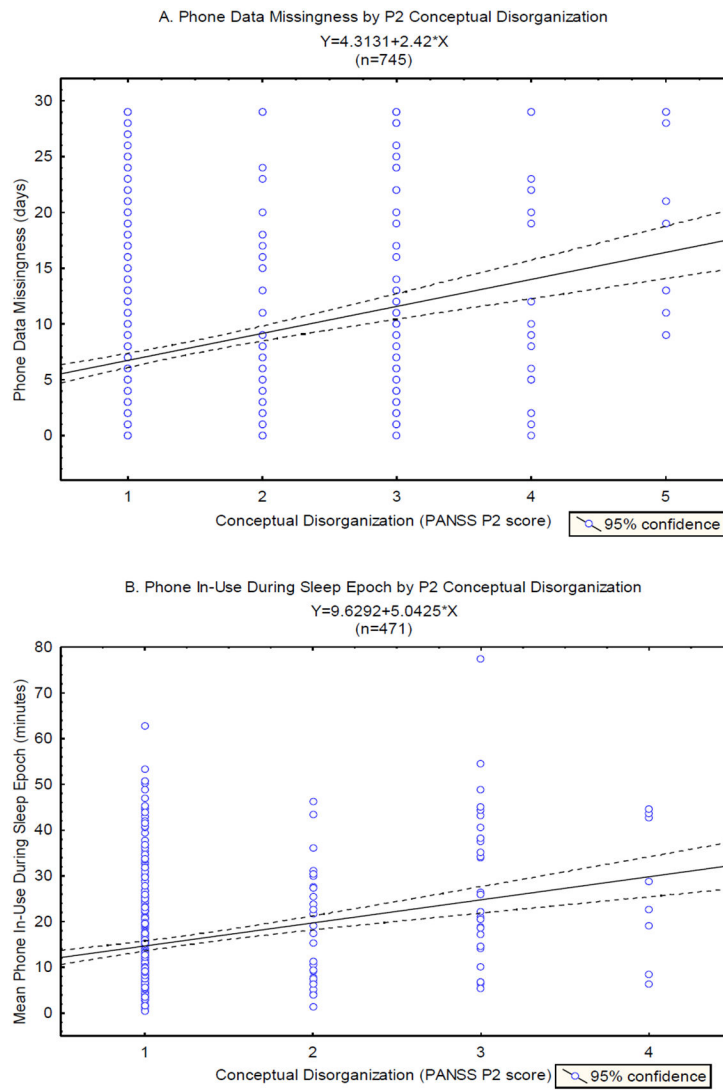
Figure 5.
A. Scatter plot of the 'phone usage data missingness' versus the PANSS P2 conceptual disorganization scores, overlaid with the linear fit line and 95% confidence inter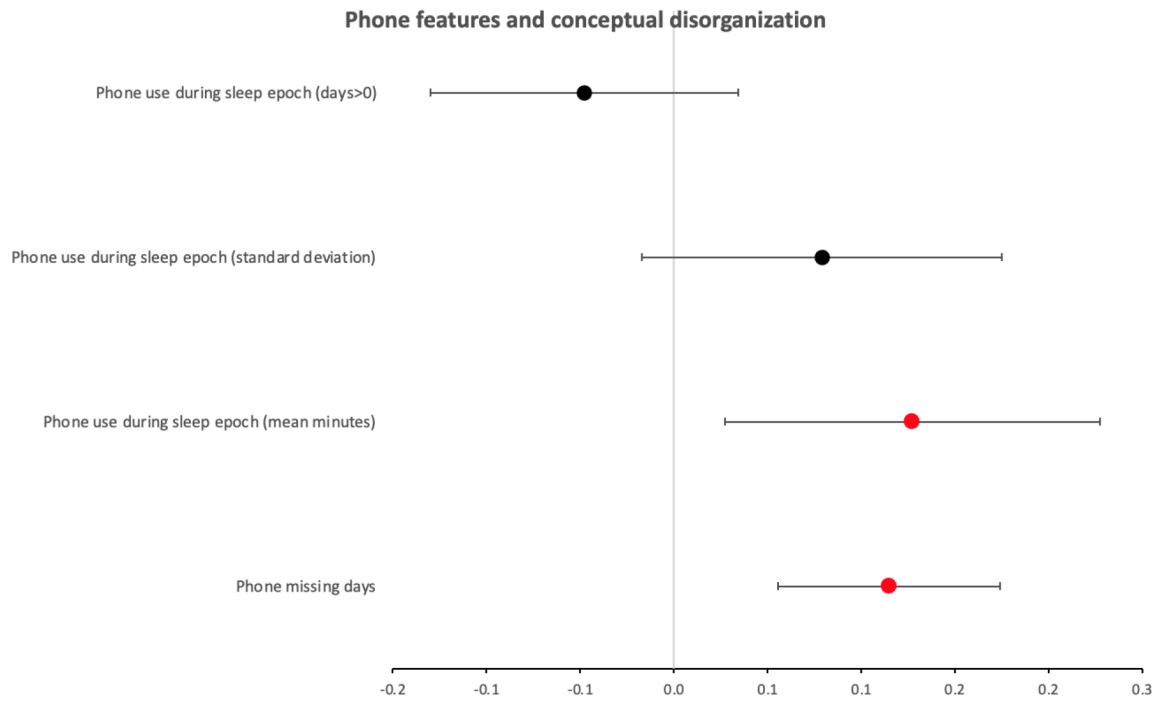val for the correlation between these variables. **B.** Scatter plot of the 'phone in-use during sleep epoch' versus PANSS P2 conceptual disorganization scores, overlaid with the linear fit line and 95% confidence interval for the correlation between these variables.

Table 1.

Description of language and phone features analyzed in the study participants

	Definition/calculation	Measured skill
Language features		
Ratio of participant words	Total words produced by participant over total words produced in interview	Spontaneity in speech and willingness to speak, verbosity
Words per sentence	Mean number of words per sentence	Sentence complexity, verbosity
Words per second	Total words produced by participant over total duration of participant's turns	Speed of speech production
Disfluencies per sentence	Mean number of disfluencies per sentence, including non-verbal and verbal edits, word repeats, and restarts	Ability to formulate sentences, fluency in conversation
Non-verbal edits per sentence	Mean number of non-verbal fillers (e.g., "uh", "ah", "hmm") per sentence	Speech fluency
Verbal edits per sentence	Mean number of verbal fillers (e.g., "like", "I mean", "you know") per sentence	Speech fluency
Word repeats per sentence	Mean number of word repeats and stutters per sentence (e.g., "I, I, I'm g-g-going out").	Speech fluency
Restarts per sentence	Mean number of restarts (e.g., "did you call—phone him?") per sentence	Speech fluency
Sequential word incoherence per sentence	Mean cosine similarity between sequential word embeddings per sentence	Semantic regularity
Pairwise word incoherence per sentence	Mean cosine similarity between embeddings of every word pair per sentence	Semantic regularity
Word uncommonness per sentence	Mean magnitude of word embeddings per sentence	Semantic regularity
Smartphone features		
Phone usage missing days	Number of days in search window for which phone data are missing (i.e, available for less than 60 minutes per day). Search window=28 days before interview day, or number of days since previous interview (if <28 days)	Study compliance, ability to maintain routine
Phone use during sleep epoch (mean)	Mean minutes of phone in-use during sleep epoch over search window. See Methods for sleep epoch definition	Sleep quality
Phone use during sleep epoch (STD)	Standard deviation (STD) of mean phone in-use during sleep epoch over search window	Seep quality
Phone use during sleep epoch (days>0)	Number of days in data window for which phone in-use during sleep was greater than zero	Sleep quality

Table 2.

Distribution of demographic variables and clinical diagnosis in the full (n=59), phone usage (n=51), and language (n=18) samples used for the analyses.

	Full Sample		Phone Sample		Language Sample	
	Count	Percent	Count	Percent	Count	Percent
Sex						
Female	37	62.7	35	68.6	7	38.9
Male	22	37.3	16	31.4	11	61.1
Race						
African American	7	11.9	7	13.7	0	0.0
American Indian	1	1.7	0	0.0	1	5.6
Asian	8	13.6	8	15.7	2	11.1
White	40	67.8	33	64.7	15	83.3
Not Reported	3	5.1	3	5.9	0	0.0
Education						
4 Years College	22	37.3	19	37.3	4	22.2
Part College	21	35.6	18	35.3	8	44.4
Graduate/Professional School	11	18.6	10	19.6	2	11.1
Highschool	4	6.8	3	5.9	3	16.7
Not Reported	1	1.7	1	2.0	1	5.6
Age						
48-51	3	5.1	2	3.9	2	11.1
40-47	3	5.1	2	3.9	1	5.6
30-39	10	16.9	8	15.7	4	22.2
24-29	19	32.2	16	31.4	4	22.2
18-23	16	27.1	15	29.4	4	22.2
Not Reported	8	13.6	8	15.7	3	16.7
Diagnosis						
Schizophrenia	5	8.5	3	5.9	2	11.1
Schizoaffective	10	16.9	10	19.6	5	27.8
Bipolar I	21	35.6	17	33.3	7	38.9
Bipolar II	12	20.3	10	19.6	4	22.2
Major Depressive Disorder	11	18.6	11	21.6	0	0.0