

Mobile RDoC: Using Smartphones to Understand the Relationship Between Auditory Verbal Hallucinations and Need for Care

Dror Ben-Zeev^{*1}, Benjamin Buck¹, Ayesha Chander¹, Rachel Brian¹, Weichen Wang², David Atkins¹, Carolyn J. Brenner¹, Trevor Cohen^{1,3}, Andrew Campbell², and Jeffrey Munson¹

¹Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; ²Department of Computer Science, Dartmouth College, Hanover, NH; ³Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA

*To whom correspondence should be addressed; BRiTE Center, Department of Psychiatry and Behavioral Sciences, 1959 NE Pacific Street, Box 356560, Seattle, WA 98195-6560, US; tel: 206-685-9655, fax: 206-543-9520, e-mail: dbenzeev@uw.edu

Objective: Auditory verbal hallucinations (AVH) are common in multiple clinical populations but also occur in individuals who are otherwise considered healthy. Adopting the National Institute of Mental Health's Research Domain Criteria (RDoC) framework, the aim of the current study was to integrate a variety of measures to evaluate whether AVH experience varies across clinical and nonclinical individuals. **Methods:** A total of 384 people with AVH from 41 US states participated in the study; 295 participants (77%) who received inpatient, outpatient, or combination treatments for AVH and 89 participants (23%) who never received care. Participants used a multi-modal smartphone data collection system to report on their AVH experiences and co-occurring psychological states multiple times daily, over 30 days. In parallel, smartphone sensors recorded their physical activity, geolocation, and calling and texting behavior continuously. **Results:** The clinical sample experienced AVH more frequently than the nonclinical group and rated their AVH as significantly louder and more powerful. They experienced more co-occurring negative affect and were more socially withdrawn, spending significantly more time at home and significantly less time near other people. Participants with a history of inpatient care also rated their AVH as infused with significantly more negative content. The groups did not differ in their physical activity or use of their smartphones for digital communication. **Conclusion:** Smartphone-assisted remote data collection revealed real-time/real-place phenomenological, affective, and behavioral differences between clinical and nonclinical samples of people who experience AVH. The study provided strong support for the application of RDoC-informed approaches in psychosis research.

Key words: schizophrenia/mobile health (mHealth)/sensing/ecological momentary assessment/mobile phones/psychosis

Introduction

Auditory verbal hallucinations (AVH) are prevalent among people diagnosed with schizophrenia but also occur among people with other psychiatric diagnoses and in individuals who are otherwise labeled "healthy."¹⁻³ Epidemiologic studies suggest AVH are not uncommon in the general population.^{4,5} Most people who report AVH do not meet diagnostic criteria for a psychotic disorder.⁶ AVH may therefore be part of a continuum of psychotic experience ranging from what would be considered "normal" to pathological.⁷⁻⁹

The National Institute of Mental Health's Research Domain Criteria (RDoC¹⁰) is a research framework that can help guide examination of AVH on a continuum.¹¹ RDoC-informed AVH research has primarily focused on neurobiological, physiological, and neurocognitive studies in people with schizophrenia-spectrum disorders.¹¹⁻¹³ RDoC-informed research that focuses on the phenomenology of AVH may further help reveal what distinguishes individuals who experience these symptoms in the context of a clinical disorder from those who experience AVH but do not require care.^{14,15} Identifying differentiating factors may help in the development of prevention strategies for subclinical populations so they do not progress in the psychotic trajectory as well as inform more targeted treatments for individuals with AVH who already require treatment.

Preliminary studies taking this approach have suggested differences that may distinguish clinical and nonclinical AVH, including perceived controllability, frequency, and valence.^{3,16–20} Johns et al¹⁴ reviewed the literature and concluded that the most significant differentiating factor was the content of AVH, with clinical individuals reporting more negative themes. Clinical and nonclinical individuals may also differ in their affective response to AVH; while a majority of clinical individuals report experiencing AVH as upsetting with moderate or severe associated anxiety and sadness,^{19,20} only a minority of subclinical individuals report significant negative affect.^{15,17}

Measuring AVH phenomenology is challenging.¹¹ For decades retrospective measures (ie, interviews, questionnaires) administered in laboratory or clinic settings were the primary method to gain insights into the subjective experience of AVH. AVH assessments typically involved asking participants to recollect (eg, “how many times did you hear voices last month?”), aggregate (eg, “on average, how distressing are the voices?”), or summarize (eg, “what do you typically do when you hear voices?”) their experience. These methods are susceptible to multiple potential sources of error, including inaccurate estimates, interpretive errors, or assessment demand characteristics.^{21,22}

Mobile technology presents new opportunities to deepen our understanding of the differences and similarities across clinical and nonclinical AVH. Smartphones have a host of embedded sensors (eg, light sensors, GPS, microphone) that can be repurposed for behavioral sensing—ie, continuous passive recording of the smartphone users’ behavior (eg, geospatial activity, physical activity, social interactions) as they go about their daily lives.^{23–25} Smartphones can also support software that facilitates Ecological Momentary Assessment (EMA)—a self-report paradigm that involves in-the-moment data capture.²⁶ In EMA studies, participants are prompted by a mobile device to complete brief self-reports on their current thoughts, affect, and immediate context repeatedly to produce multiple “snapshots” over a given period. Smartphones have been used successfully for both EMA^{27,28} and behavioral sensing^{29,30} among participants who experience AVH.

Adopting the RDoC framework’s dimensional approach to psychopathology research, the objective of the current exploratory study was to integrate a variety of measures to evaluate whether AVH experience and related behaviors vary across clinical and nonclinical individuals. To do so, we combined EMA self-reporting methodology and behavioral sensing technology into an integrative smartphone data collection system³¹ that enabled us to remotely capture participants’ phenomenological AVH experiences and co-occurring psychological states and behavior over 30 days.

Methods

Participants

A total of 384 individuals with AVH completed data collection. Participants were recruited remotely online or via community-based strategies in Seattle. Participants met inclusion criteria if they were (1) 18 years or older; (2) an English speaker; (3) able to use a smartphone and (4) reported experiencing AVH at least once weekly. Participants were excluded if they (1) did not live in the United States, (2) had already participated in the study, or (3) were unavailable for the 30 days of data collection. To allow for remote participation, those recruited online were also required to own an Android smartphone with an active data plan. To ensure representation of a broader range of individuals (ie, people with limited resources or technology familiarity), participants engaged through community efforts were included if they did not own a smartphone.

Procedures

The study was approved by the Institutional Review Boards of the University of Washington and Dartmouth College. Online recruitment was conducted with the aid of Google Ads. These online ads present postings to users based on the extent to which their search terms match pre-selected overtly clinical (eg, *schizophrenia*, *bipolar*, *hearing voices*), colloquial (eg, *talking to ghosts*, *am I crazy*, *stress relief*), and related (ie, generated by the Google Ads “broad match” algorithm) keywords that may pertain to experiencing AVH. Individuals who clicked on ads were directed to the study website, which included an infographic and videos explaining the project, and a link to the consent form. Through the website, participants could verify their phone number and email, complete a competency screening questionnaire, provide informed consent, complete baseline assessments, and download the study application.

Community participants were recruited from the Seattle area using flyers, practitioner referrals, and participant snowball referrals. Research staff spoke to potential participants by phone to ask pre-screening questions and share additional study details. If candidates were eligible and interested, staff scheduled an in-person study visit. During that visit, research staff directed participants to the study website and assisted them with the same procedures accessed by those recruited online. After the participant completed all assessments, staff oriented participants to a smartphone they provided with the data collection application installed.

All participants were instructed to carry the smartphone and respond to prompts for 30 days. Participants could contact study staff directly to ask questions or receive technical assistance. After 30 days, the application stopped sending data to the research team. All participants were offered \$75 for participating.

Community participants were given the option to keep the study device instead of receiving payment.

Measures

Baseline Measures. The Hamilton Program for Schizophrenia Voices Questionnaire (HPSVQ³²), a 13-item self-report measure was used to AVH severity. Depressive symptoms were assessed with the Patient Health Questionnaire (PHQ-9³²) a 9-item assessment of depressive symptoms that is summed for an overall score. General psychopathology was assessed with the Symptom Checklist-9 (SCL-9³³), a 9-item brief version of the 90-item Symptom Checklist, which assesses a range of general symptoms including somatization, worry, and anger using the average score across items. The Sheehan Disability Scale (SDS³⁴) was used to

assess symptom-related impairments in work/school activities, family relationships, and social functioning. Participants were asked to report on whether they had ever sought treatment specifically for their AVH. Participants could endorse multiple responses but were classified into one of the followings of care groups for analyses, based on their highest treatment intensity for AVH: no treatment, outpatient treatment, or inpatient treatment. Participant demographics are presented in [table 1](#).

Mobile Data Collection

The integrated smartphone data collection application is an updated version of a system validated and deployed in our previous research,^{29,31,33} modified for the needs of this study.

Table 1. Participant Characteristics

		Inpatient Treatment (N = 147)	Other Treatment (N = 148)	No Treatment (N = 89)	
Age		39.2 (10.7)	40.6 (11.7)	42.4 (12.1)	
Gender	Female	75 (51.0%)	72 (48.6%)	45 (50.6%)	
	Male	66 (44.9%)	70 (47.3%)	40 (44.9%)	
	Transgender: MTF	1 (0.7%)	4 (2.7%)	2 (2.2%)	
	Transgender: FTM	3 (2.0%)	2 (1.4%)	0 (0.0%)	
	Other	2 (1.4%)	0 (0.0%)	2 (2.2%)	
Race	White	99 (67.3%)	82 (55.4%)	57 (64.0%)	
	Black or African American	22 (15.0%)	46 (31.1%)	12 (13.5%)	
	American Indian or Alaskan Native	1 (0.7%)	2 (1.4%)	3 (3.4%)	
	Asian	3 (2.0%)	3 (2.0%)	1 (1.1%)	
	More than one race	21 (14.3%)	13 (8.8%)	15 (16.9%)	
	Missing / Declined	1 (0.7%)	2 (1.4%)	1 (1.1%)	
	Ethnicity	Hispanic / Latino	25 (17.0%)	22 (14.9%)	11 (12.4%)
		Not Hispanic / Latino	120 (81.6%)	126 (85.1%)	78 (87.6%)
		Missing / Declined	2 (1.4%)	0 (0.0%)	0 (0.0%)
	Marital status	Never married	66 (44.9%)	74 (50.0%)	32 (36.0%)
Living with someone		16 (10.9%)	16 (10.8%)	9 (10.1%)	
Married		19 (12.9%)	18 (12.2%)	12 (13.5%)	
Separated		14 (9.5%)	12 (8.1%)	13 (14.6%)	
Divorced		26 (17.7%)	22 (14.9%)	20 (22.5%)	
Widowed		6 (4.1%)	6 (4.1%)	3 (3.4%)	
Sexual orientation	Heterosexual or straight	104 (70.7%)	118 (79.7%)	73 (82.0%)	
	Gay or lesbian	14 (9.5%)	10 (6.8%)	5 (5.6%)	
	Bisexual	21 (14.3%)	14 (9.5%)	10 (11.2%)	
	Other	7 (4.8%)	6 (4.1%)	1 (1.1%)	
	Missing / Declined	1 (0.7%)	0 (0.0%)	0 (0.0%)	
	Employment status	Unemployed	110 (74.8%)	111 (75.0%)	60 (67.4%)
Working part-time		23 (15.6%)	22 (14.9%)	15 (16.9%)	
Working full-time		14 (9.5%)	14 (9.5%)	14 (15.7%)	
Missing / Declined		0 (0.0%)	1 (0.7%)	0 (0.0%)	
Living situation	Independent/Living on my own	66 (44.9%)	76 (51.4%)	33 (37.1%)	
	Living with family	53 (36.1%)	48 (32.4%)	31 (34.8%)	
	Assisted/supported living	14 (9.5%)	8 (5.4%)	6 (6.7%)	
	Substance treatment institution	0 (0.0%)	2 (1.4%)	1 (1.1%)	
	Homeless	14 (9.5%)	14 (9.5%)	18 (20.2%)	

EMA

The application prompted participants to complete a 12-item self-report assessment 4 times daily between 9 AM and 9 PM. The first item asked participants to report whether they were experiencing AVH (“Are you experiencing VOICES right now?” *Yes/No*) and continued to the subsequent questions if participants indicated that they were experiencing AVH. If AVH was endorsed, participants were asked additional questions about their hallucinations using a 4-point scale (*1-Not at all; 2-A little; 3-Moderately; 4-Extremely*) unless response options were categorical in nature. Participants had the option to *skip* items they did not want to respond to. Questions included (1) 3 ratings of *AVH experience* (negativity: “How NEGATIVE is the content of the voices?” rated 1 to 4, loudness: “How LOUD are the voices?” rated 1 to 4, safety: “How SAFE do you feel right now?” rated 1 to 4), (2) 3 ratings of *AVH appraisals* (control: “How much CONTROL do you have over the voices?” rated 1 to 4, power: “How much POWER do the voices have?” rated 1 to 4, and externality: “Where are the voices COMING FROM?” rated categorically: *Inside my head; Outside my head; Both; Not Sure*), and (3) 3 ratings of *affect* (distress: “How DISTRESSED do you feel right now?” rated 1 to 4, anxiety: “How ANXIOUS do you feel right now?” rated 1 to 4, and sadness: “How SAD do you feel right now?” rated 1 to 4). Given the strong positive correlations among these 3 affect ratings (average $r = .68$), a composite negative affect score was calculated using the mean of these items. Higher ratings reflected a more negative experience except for the safety and control questions where higher values indicated a greater sense of safety and control. In addition to these ratings, participants responded to (4) 2 *Yes/No* items regarding characteristics of context (alone: “Are you ALONE right now?,” and in public, “Are you in a PUBLIC PLACE right now?”).

Behavioral Sensing and Device Use

Geospatial Activity. Using GPS, Wi-Fi, and cellular tower location services, time-stamped estimated locations were sampled and recorded every 10 minutes on the device. Estimates of distance traveled were calculated along with time at locations in which participants spent longer periods. For the current analysis, we included the number of unique locations visited and time spent in one’s primary location.

Physical Activity. The study application used Google Activity Recognition API (Application Programming Interface) in order to log physical activity. This API uses a dynamic algorithm to determine activity using device sensors. Activities were continuously assessed on the device while the user remained active. If the device was still for prolonged periods of time the sampling was only

captured once the device shifted from “still” to “active.” For the present analysis, we examined total physical activity, which included 3 variables representing time spent on foot, in a vehicle, or on a bicycle.

Speech Frequency and Duration. Using the smartphone microphone to capture ambient sound, the study application applied a speech detection algorithm to assess when human speech was nearby. The microphone sampled every 3 minutes to assess surrounding sound in proximity of the device. To protect privacy, no raw audio was captured on the device as part of the sensing system, but instead would classify the data in the moment and would only record the presence of speech. For the current analysis we included the total duration of speech (minutes per day) and the count of discrete periods of speech detected.

Phone Calls and SMS Messages. The study application logged SMS text message exchanges (sent and received), as well as number and duration of phone calls. To protect privacy, no written or audio content from these calls or messages were recorded.

Data Analytic Plan

Models examining group differences among participant levels of care for AVH (ie, no treatment, outpatient-level, inpatient-level) were constructed with 2 contrast variables, one which compared the inpatient group to the no treatment and another that compared the outpatient treatment group to the no treatment group. Baseline group differences were examined using linear regression models with these 2 contrast variables as predictors. EMA ratings were analyzed using generalized mixed effects models (R packages: *lme4*³⁵, *glmmTMB*³⁶) with the repeated EMA’s nested within individual, the 2 group contrasts treated as fixed effects, and the appropriate distribution family and link function for each dependent measure. Behavioral sensing data are reported based on measures aggregated across each day of data collection and were also analyzed using generalized mixed effect models. A number of these measures (eg, number of calls and SMS messages) exhibited overdispersion and excess zeroes. In these cases, we used a mixed-effects negative binomial hurdle model, which separates the analysis into 2 parts: (1) a logit model for zero versus non-zero values, and (2) a truncated-at-zero negative binomial for positive counts.

Results

A total of 384 individuals (305 recruited online; 79 recruited in the local community) from 41 US states completed data collection. Online, 522 people who viewed the study website and passed the screener to enroll in the study never downloaded the app and did not engage in data collection. From the community, 62 people passed

the screener to enroll in the study never downloaded the app and did not engage in data collection. Participants received 120 prompts to complete EMA questionnaires and completed on average 72.2 (60% response rate) over the data collection period. The study sample's mean age was 40.5 years old (SD = 11.5) and half the sample was female. Eighty-nine participants (23.2%) reported never receiving care for AVH. Of the 295 participants who received treatment (76.8%), 90% reported receiving individual therapy, 50% received inpatient care, 33% partial hospitalization, 31% group therapy, 23% alcohol/drug rehabilitation, 15% residential treatment services, 12% online treatments, and 6% telepsychiatry (respondents could endorse more than one option). Participants who reported receiving inpatient treatment reported an average of 3.6 (SD = 1.4) different treatment types, those with only outpatient treatments reported 1.6 (SD = 0.9).

Baseline Measures

Treatment intensity groups significantly differed on multiple measures of psychopathology. Specifically, participants with inpatient treatment history showed significantly higher scores on the SCL, PHQ-9, HPSVQ, and SDS compared to the non-clinical group. Similarly, participants with a history of outpatient-only treatment history also showed significantly higher scores on the PHQ-9, HPSVQ, and SDS compared to those with no treatment history (table 2).

EMA

Across the entire sample, a total of 27 731 EMA self-reports were collected; an average of 72.2 (SD = 25.8) per participant. Results indicated that both the inpatient-level group ($\beta = .57$, SE = .26, $Z = 2.19$, $P < .05$, odds ratio = 1.77) and outpatient-level group ($\beta = .89$, SE = .26, $Z = 3.44$, $P < .001$, odds ratio = 2.44) were more likely than the no treatment group (intercept = -1.38,

SE = .21, $Z = -6.69$, $P < .001$, odds ratio = 0.25) to report AVH occurring during EMA reports. Overall, the no treatment group endorsed AVH in 27.6% of EMAs, the inpatient-level group in 33.8%, and the outpatient-level group in 37.3%.

For AVH experience, appraisal, affect variables, we used mixed-effects models with the 2 contrast variables to compare the experience of AVH between the inpatient-level and outpatient-level groups and the no treatment group. The inpatient-level group rated their experience of AVH as significantly more negative than the no-treatment group. The inpatient-level group rated their AVH as more negative ($\beta = .31$, $Z = 2.97$, $P < .01$), louder ($\beta = .21$, $Z = 2.46$, $P < .05$), and more powerful ($\beta = .32$, $Z = 3.12$, $P < .01$). The inpatient group also reported higher levels of negative affect ($\beta = .30$, $Z = 3.06$, $P < .01$) when endorsing AVH. The outpatient-level group rated their AVH as louder ($\beta = .17$, $Z = 2.03$, $P < .05$), and more powerful ($\beta = .25$, $Z = 2.51$, $P < .05$) than the no-treatment group, but did not differ in their ratings of how negative the voices were and their own feelings of negative affect (table 3). There were no group differences in the ratings of how safe participants felt, how much control they felt they had over voices, or the likelihood of being alone or in public.

Behavioral Sensing and Device Use

Location. There were significant differences between the groups with regard to number of locations visited (table 4), as both inpatient-level ($M = 2.22$ locations, SD = 1.50, $N = 3701$; $\beta = -.16$, $Z = -2.77$, $P < .01$) and outpatient-level groups ($M = 2.27$ locations, SD = 1.64, $N = 3818$, $\beta = -0.14$, $Z = -2.49$, $P < .05$) visited significantly fewer locations than the no treatment group ($M = 2.61$ locations, SD = 1.77, $N = 2211$).

Groups also differed in the time they spent in their primary location, as the inpatient-level group ($M = 6.29$

Table 2. Group Differences in Baseline Psychopathology Measures

Variable	Tx Intensity	M	SD	N	R ²	β	SE	t
SCL score	No treatment	2.11	0.82	88	0.090	2.11	0.09	2.45*
	Inpatient	2.39	0.90	145				
	Outpatient	2.14	0.82	146				
PHQ9 total	No treatment	15.78	7.06	87	0.034	15.78	0.70	0.20
	Inpatient	19.00	6.48	146				
	Outpatient	17.63	6.37	145				
HPSVQ total	No treatment	17.11	6.57	88	0.023	17.11	0.67	3.61**
	Inpatient	22.24	6.36	147				
	Outpatient	20.90	6.03	147				
SDS total	No treatment	15.33	9.76	89	0.081	15.33	0.93	4.47**
	Inpatient	21.86	8.69	147				
	Outpatient	20.94	8.27	148				

Note: * $P < .05$; ** $P < .001$.

Table 3. Mixed Effects Models Comparing AVH Experience by Treatment Intensity

Predictors	How negative are the voices?			How loud are the voices?			How safe do you feel?			How much control do you have over the voices?			How powerful are the voices?			How distressed, anxious, sad do you feel? (avg)		
	β	SE	t	β	SE	t	β	SE	t	β	SE	t	β	SE	t	β	SE	t
(Intercept)	1.17***	0.08	14.03	1.36***	0.07	19.64	1.94***	0.07	26.34	1.03***	0.08	12.62	1.07***	0.08	13.06	1.26***	0.08	15.93
Inpatient	0.31**	0.10	2.97	0.21*	0.09	2.46	-0.10	0.09	-1.05	0.08	0.10	0.74	0.32**	0.10	3.12	0.30**	0.10	3.06
Outpatient	0.16	0.10	1.50	0.17*	0.09	2.03	0.08	0.09	0.87	0.08	0.10	0.81	0.25*	0.10	2.51	0.15	0.10	1.52
Random Effects																		
σ^2		0.60			0.66			0.66			0.62			0.64			0.41	
τ_{00}		0.50			0.37 _{id}			0.37 _{id}			0.47 _{id}			0.47 _{id}			0.47 _{id}	
ICC		0.46			0.38			0.36			0.43			0.42			0.53	
N																		
Observations																		
Marginal R ²			0.011/0.463			0.007/0.388						0.001/0.430						
Conditional R ²																		0.014/0.541

Note: *P < .05, **P < .01, ***P < .001.

hours, SD = 5.72, N = 3570, $\beta = -.41$, Z = -4.02, P < .001) and outpatient-level group (M = 6.63 hours, SD = 5.96, N = 3708, $\beta = -.34$, Z = -3.39, P < .001) spent less time away from the primary location than did the no treatment group (M = 8.30 hours, SD = 6.52, N = 2161). There were no group significant differences in average distance traveled for days in which distance traveled was greater than zero.

Speech Duration. Groups differed in the number of periods proximal to speech (table 4), as both the inpatient-level (M = 24.9, SD = 18.5, N = 3727, $\beta = -0.46$, Z = -2.25, P < .05) and outpatient-level group (M = 23.5, SD = 17.7, N = 3860, $\beta = -0.62$, Z = -3.03, P < .01) had significantly fewer periods near speech than the no treatment group (M = 29.8, SD = 20.6, N = 2237). For the duration of speech (minutes/day) the outpatient-level group (M = 249.7, SD = 223.6) showed fewer minutes than the no treatment group (M = 299.1, SD = 244.9), but this fell just short of significance, $\beta = -1.42$, Z = -1.88, P = .061.

Phone Calls and SMS Messages. Groups did not significantly differ in the likelihood of making and receiving any calls or texts (zero-inflation portion of the models), nor in the number of calls and texts made (continuous portion of the models).

Physical Activity. Groups did not significantly differ in the likelihood of recording any physical activity (zero-inflation portion of the models), nor in the amount of time spent engaged in physical activity (continuous portion of the models).

Discussion

To our knowledge, this is the first study to use multi-modal smartphone data collection techniques to capture the real-time/real-place experience and co-occurring behaviors associated with hearing voices in clinical and nonclinical samples of people with AVH. We found several key differences between these groups. First, groups differed with regard to AVH phenomenology. The clinical group experienced voices that they reported to be more frequent, louder, and more powerful than those experienced by the nonclinical group. People who required inpatient treatment rated their AVH as having significantly more negative content. Second, level of clinical care was associated with affective response to AVH. The clinical group had more negative affect when hearing voices than the nonclinical group. Finally, history of clinical care was associated with current behavior in the context of AVH. The clinical group was more socially withdrawn than the nonclinical sample; individuals with history of clinical care spent more time at home (as determined by smartphone geolocation) and spent less time near other people (as determined by speech detection software). The groups

Table 4. Behavioral Sensing Data by Treatment Intensity: Location and Speech Measures

<i>Predictors</i>	Number of Locations Visited ^a			Hours Spent Away From Primary Location ^b			Duration of Speech (Min/Day) ^b			Number of Periods of Speech ^b		
	IRR	SE	Z	β	SE	Z	β	SE	Z	β	SE	Z
(Intercept)	2.40***	0.04	19.98	2.84***	0.08	35.56	16.14***	0.60	27.06	5.21***	0.16	32.23
Inpatient	0.86**	0.06	-2.77	-0.41***	0.10	-4.02	-1.06	0.76	-1.39	-0.46*	0.21	-2.25
Outpatient	0.87*	0.06	-2.46	-0.34***	0.10	-3.39	-1.42	0.75	-1.88	-0.62**	0.20	-3.03
Random effects												
σ^2		0.38			22.43			21198.56			150.53	
τ_{00}		0.15 _{id}			0.52 _{id}			29.87 _{id}			2.21 _{id}	
ICC		0.28			0.02			0.00			0.01	
N		376 _{id}			376 _{id}			376 _{id}			376 _{id}	
Observations		9824			9439			9824			9824	
Marginal R ² / Conditional R ²		0.007/ 0.284			0.001/ 0.024			0.000/ 0.001			0.000/ 0.015	

Note: IRR, Incidence rate ratio.

^aFamily = Poisson.

^bFamily = Gaussian, link = square root.

* $P < .05$; ** $P < .01$; *** $P < .001$.

did not differ in their ratings of how safe participants felt, how much control they felt they had over voices, or the likelihood of being alone or in public. The differences we found between clinical and nonclinical groups of people who experience AVH provide further support for the utility of applying dimensional approaches to psychopathology, and the RDoC framework to hallucinations research specifically.¹⁰⁻¹²

Our results provide preliminary evidence for a number of characteristics that may determine need for care. The findings suggesting that clinical AVH are more frequent and more negatively themed are consistent with prior studies (see Baumeister et al³⁷ for a review) comparing clinical and nonclinical AVH on laboratory-based measures. One differing finding concerned the result that individuals with clinical AVH perceived their voices to be louder than those with nonclinical AVH. Most studies utilizing laboratory-based measures have suggested that clinical and nonclinical AVH do not differ in this regard.³ It is possible that the real-time, real-place approach was granular enough to capture differences that are missed in retrospective recall. From a causal perspective, it is possible that louder AVH are particularly distracting, making it difficult to attend to other important internal processes (eg, planning one's day, recalling verbal information) or to focus on external stimuli (ie, conversations with others). These difficulties may lead to greater distress or dysfunction, motivating people to seek care.

The study findings also support differences between clinical and nonclinical groups' appraisal and responses to AVH. The cognitive model of psychosis^{38,39} suggests that appraisals determine the impact of AVH on emotions and behavior; individuals who appraise their voices as more important, powerful, and difficult-to-control are thus more susceptible to negative affect and

dysfunction. Our results demonstrated that individuals with clinical AVH appraised their symptoms as more powerful than those with nonclinical AVH. These findings may be consistent with Birchwood and colleagues' application of social rank theory to psychosis modeling and empirical findings suggesting one's subjective sense of subordination to voices is closely associated with their general sense of social powerlessness and marginalization more broadly.^{40,41} Those with inpatient treatment history also experienced higher distress, and those in both clinical groups reported behavioral disruption as evidenced by greater isolation, more time spent at home, and fewer locations visited. While this appears to support individual components of the cognitive model of psychosis, questions remain about causal relationships. On one hand, louder and more distressing voices may cause those with clinical AVH to isolate; on the other hand, individuals who are more socially isolated may be more likely to attend to their voices and over-emphasize their importance, thus increasing their AVH-related distress.

The factors that distinguished clinical and nonclinical samples suggest potential targets for intervention. Several interventions—eg, social skills training,⁴² behavioral activation⁴³—are designed to increase social interaction for individuals with serious mental illnesses; these results suggest that aspects of these strategies could be particularly apt for transdiagnostic AVH. Further, our results suggest that cognitive interventions could specifically target appraisals of AVH as powerful. Given the fact that participants in the present study carried and responded to a digital assessment tool for 1 month, it is possible that an intervention could be delivered via the same modality, particularly given high rates of smartphone ownership and interest in mHealth⁴⁴

among individuals with psychosis or at risk for it (eg, refs.^{28,45}).

This exploratory study has several strengths. Our use of mobile data collection techniques enabled us to record dimensions of AVH in a manner less susceptible to recall biases or reporting inaccuracies. Additionally, our study's remote recruitment and assessment procedures may have allowed for greater representation of the heterogeneity of this population in research. These online recruitment, screening, enrollment, and assessment methods allowed us to engage a varied and geographically dispersed sample, efficiently reaching 41 US states. Technology-assisted remote techniques have been used successfully with other clinical populations⁴⁶ and this study demonstrated their feasibility with people who experience psychotic symptoms.

The study has limitations. First, our study was observational in nature. It cannot be known whether group differences in AVH experience are the cause of need for treatment, the result of it, or whether a third factor accounts for both. Second, mobile sensing captures digital traces that we interpreted as representative of behavior but such inferences should be made with caution. For example, automated detection of human speech typically suggests that the individual carrying the device was in a social environment. However, they might have also been in close proximity to a television set depicting people speaking, an event that may be classified erroneously by our study software. Third, mobile prompts to complete EMA measures could have impacted respondents' thoughts, feelings, and behavior during the data collection period. This could be attributable to a Hawthorne Effect or to the potential benefits of ongoing tracking of one's own clinical symptoms (eg, more frequent self-reflection, increased insight into factors affecting change). Fourth, our EMA item assessing AVH valence focused solely on the intensity of AVH negativity. While negative content has been identified as one of the most significant differentiating factors between clinical and nonclinical groups in past research, including an EMA item evaluating positivity would have allowed us to examine the balance between these 2 forms of AVH content, as they pertain to need for care.¹⁴ Finally, in the study "need for care" was operationalized as self-reported history of receiving care. While the 2 are certainly closely related, it is possible that some individuals who required care never sought it, or that some of those that sought care were unsuccessful in receiving it (eg, being waitlisted for treatment at the clinic).

Conclusion

A growing body of literature suggests AVH experiences are diverse. A continuum may separate forms of AVH that lead to distress, impairment, and need for treatment from

perceptual experiences that do not require care. The present study helped identify several phenomenological and behavioral variables that may determine where one falls on this continuum. Continued research is needed to translate these findings into clinically useful interventions that may make the difference between episodes that are experienced as debilitating and more successful coping and resilience.

Funding

National Institute of Mental Health (R01MH112641).

Acknowledgments

D.B-Z. has consulted to Trusst Health, eQuility, and Otsuka Pharmaceuticals Ltd. He has an intervention content licensing agreement with Pear Therapeutics and has a financial interest in FOCUS technology.

References

1. Larøi F, Sommer IE, Blom JD, *et al*. The characteristic features of auditory verbal hallucinations in clinical and nonclinical groups: state-of-the-art overview and future directions. *Schizophr Bull*. 2012;38:724–733.
2. Sommer IE, Daalman K, Rietkerk T, *et al*. Healthy individuals with auditory verbal hallucinations; who are they? Psychiatric assessments of a selected sample of 103 subjects. *Schizophr Bull*. 2010;36:633–641.
3. Daalman K, Boks MP, Diederer KM, *et al*. The same or different? A phenomenological comparison of auditory verbal hallucinations in healthy and psychotic individuals. *J Clin Psychiatry*. 2011;72:320–325.
4. Tien AY. Distributions of hallucinations in the population. *Soc Psychiatry Psychiatr Epidemiol*. 1991;26:287–292.
5. Scott J, Chant D, Andrews G, McGrath J. Psychotic-like experiences in the general community: the correlates of CIDI psychosis screen items in an Australian sample. *Psychol Med*. 2006;36:231–238.
6. Johns LC, Nazroo JY, Bebbington P, Kuipers E. Occurrence of hallucinatory experiences in a community sample and ethnic variations. *Br J Psychiatry*. 2002;180:174–178.
7. Van Os J, Hanssen M, Bijl RV, Ravelli A. Strauss (1969) revisited: a psychosis continuum in the general population? *Schizophr Res*. 2000;41:8.
8. Johns LC, van Os J. The continuity of psychotic experiences in the general population. *Clin Psychol Rev*. 2001;21:1125–1141.
9. van Os J, Linscott RJ, Myin-Germeys I, Delespaul P, Krabbendam L. A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychol Med*. 2009;39:179–195.
10. Cuthbert BN; NIMH RDoC Workgroup. The RDoC framework: continuing commentary. *World Psychiatry*. 2014;13:196–197.
11. Ford JM. Studying auditory verbal hallucinations using the RDoC framework. *Psychophysiology*. 2016;53:298–304.
12. Ford JM, Morris SE, Hoffman RE, *et al*. Studying hallucinations within the NIMH RDoC framework. *Schizophr Bull*. 2014;40(Suppl 4):S295–S304.

13. Badcock JC, Hugdahl K. A synthesis of evidence on inhibitory control and auditory hallucinations based on the Research Domain Criteria (RDoC) framework. *Front Hum Neurosci*. 2014;8:180. doi:10.3389/fnhum.2014.00180
14. Johns LC, Kompus K, Connell M, et al. Auditory verbal hallucinations in persons with and without a need for care. *Schizophr Bull*. 2014;40(Suppl 4):S255–S264.
15. Romme MAJ, Escher ADM a. c. Hearing voices. *Schizophr Bull*. 1989;15:209–216.
16. Honig A, Romme MA, Ensink BJ, Escher SD, Pennings MH, deVries MW. Auditory hallucinations: a comparison between patients and nonpatients. *J Nerv Ment Dis*. 1998;186:646–651.
17. Lawrence C, Jones J, Cooper M. Hearing voices in a non-psychiatric population. *Behav Cogn Psychother*. 2010;38:363–373.
18. Leudar I, Thomas P, McNally D, Glinski A. What voices can do with words: pragmatics of verbal hallucinations. *Psychol Med*. 1997;27:885–898.
19. Freeman D, Garety PA. Connecting neurosis and psychosis: the direct influence of emotion on delusions and hallucinations. *Behav Res Ther*. 2003;41:923–947.
20. Hoffman RE, Varanko M, Gilmore J, Mishara AL. Experiential features used by patients with schizophrenia to differentiate ‘voices’ from ordinary verbal thought. *Psychol Med*. 2008;38:1167–1176.
21. Khan A, Yavorsky WC, Liechti S, et al. Assessing the sources of unreliability (rater, subject, time-point) in a failed clinical trial using items of the Positive and Negative Syndrome Scale (PANSS). *J Clin Psychopharmacol*. 2013;33:109–117.
22. Ben-Zeev D, McHugo GJ, Xie H, Dobbins K, Young MA. Comparing retrospective reports to real-time/real-place mobile assessments in individuals with schizophrenia and a nonclinical comparison group. *Schizophr Bull*. 2012;38:396–404.
23. Gurrin C, Qiu Z, Hughes M, et al. The smartphone as a platform for wearable cameras in health research. *Am J Prev Med*. 2013;44:308–313.
24. del Rosario MB, Redmond SJ, Lovell NH. Tracking the evolution of smartphone sensing for monitoring human movement. *Sensors*. 2015;15:18901–18933.
25. Rabbi M, Ali S, Choudhury T, Berke E. Passive and in-situ assessment of mental and physical well-being using mobile sensors. In: *Proceedings of ACM International Conference on Ubiquitous Computing*. 2011;2011:385–394. doi:10.1145/2030112.2030164
26. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. In: *Annual Review of Clinical Psychology* 2008;4:1–32. doi:10.1146/annurev.clinpsy.3.022806.091415.
27. Palmier-Claus JE, Ainsworth J, Machin M, et al. The feasibility and validity of ambulatory self-report of psychotic symptoms using a smartphone software application. *BMC Psychiatry*. 2012;12:172.
28. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. *Schizophr Bull*. 2014;40:1244–1253.
29. Ben-Zeev D, Wang R, Abdullah S, et al. Mobile behavioral sensing for outpatients and inpatients with schizophrenia. *Psychiatr Serv*. 2016;67:558–561.
30. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela JP. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology*. 2018;43:1660–1666.
31. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. *Psychiatr Rehabil J*. 2015;38:218–226.
32. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann*. 2002;32:509–515.
33. Ben-Zeev D, Brian R, Wang R, et al. CrossCheck: integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. *Psychiatr Rehabil J*. 2017;40:266–275.
34. Sheehan DV. *The Anxiety Disease*. New York, NY: Scribner; 1983.
35. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2014. arXiv:1406.5823 [stat.CO].
36. Brooks ME, Kristensen K, van Benthem KJ, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J*. 2017;9:378–400. doi:10.32614/rj-2017-066
37. Baumeister D, Sedgwick O, Howes O, Peters E. Auditory verbal hallucinations and continuum models of psychosis: a systematic review of the healthy voice-hearer literature. *Clin Psychol Rev*. 2017;51:125–141.
38. Garety PA, Kuipers E, Fowler D, Freeman D, Bebbington PE. A cognitive model of the positive symptoms of psychosis. *Psychol Med*. 2001;31:189–195.
39. Mawson A, Cohen K, Berry K. Reviewing evidence for the cognitive model of auditory hallucinations: the relationship between cognitive voice appraisals and distress during psychosis. *Clin Psychol Rev*. 2010;30:248–258.
40. Birchwood M, Gilbert P, Gilbert J, et al. Interpersonal and role-related schema influence the relationship with the dominant ‘voice’ in schizophrenia: a comparison of three models. *Psychol Med*. 2004;34:1571–1580.
41. Birchwood M, Meaden A, Trower P, Gilbert P, Plaistow J. The power and omnipotence of voices: subordination and entrapment by voices and significant others. *Psychol Med*. 2000;30:337–344.
42. Kurtz MM, Mueser KT. A meta-analysis of controlled research on social skills training for schizophrenia. *J Consult Clin Psychol*. 2008;76:491–504.
43. Cuijpers P, van Straten A, Warmerdam L. Behavioral activation treatments of depression: a meta-analysis. *Clin Psychol Rev*. 2007;27:318–326.
44. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile phone ownership and endorsement of “mHealth” among people with psychosis: a meta-analysis of cross-sectional studies. *Schizophr Bull*. 2016;42:448–455. doi:10.1093/schbul/sbv132
45. Bell IH, Rossell SL, Farhall J, et al. Pilot randomised controlled trial of a brief coping-focused intervention for hearing voices blended with smartphone-based ecological momentary assessment and intervention (SAVVy): feasibility, acceptability and preliminary clinical outcomes. *Schizophr Res*. 2020;216:479–487.
46. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. *BMJ Innov*. 2016;2:14–21.
47. U.S. Census Bureau. *Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019*; 2019.