

Consistency and reliability of the auditory vocal hallucination rating scale (AVHRS)

Received 28 September 2011; Revised 21 January 2012; Accepted 29 January 2012; First published online 13 March 2012

Key words: Auditory hallucinations, interview, psychometrics, research.

Dear Editor

Assessing characteristics of auditory vocal hallucinations (AVH) is a key issue in psychiatric clinical practice and research. Recently, a review of assessment tools for AVH has been published (Ratcliff *et al.* 2011), describing two structured interviews and eight self-report instruments. Meanwhile, the auditory vocal hallucinations rating scale (AVHRS) (Jenner & Van de Willige, 2002) developed since 1996 has been used in adolescent and adult patient research (Van de Willige *et al.* 1996; Jenner & Van de Willige, 2001) and in research on non-clinical children (Bartels-Velthuis *et al.* 2010, 2011, 2012). We examined the consistency, the inter-rater reliability and the discriminative power of the AVHRS.

Methods

Instruments

Development of the AVHRS

At the voices outpatient department (VOPD), the auditory hallucination rating scale (AHRS) was developed to examine the effectiveness of therapy (Van de Willige *et al.* 1996). Items were derived from clinical practice, from symptom characteristics formulated in the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (Wing *et al.* 1990) and from DSM-IV (American Psychiatric Association, 1994).

In a population-based survey among 7- and 8-year-old children assessing prevalence and characteristics of auditory vocal hallucinations (AVH) (Bartels-Velthuis *et al.* 2010), the initial AHRS was combined with items of the auditory hallucinations subscale (AHS) of the psychotic symptom rating scale PSYRATS (Haddock *et al.* 1999), resulting in the auditory vocal hallucinations rating scale (AVHRS) (Jenner & Van de Willige, 2002). In addition, three items were

added: (i) voices talking separately or simultaneously, guided by the observations that patients' suffering increases when voices are talking simultaneously; (ii) hypnagogic and/or hypnopompic hallucinations (HHH), which might be regarded as non-pathological and not uncommon in accompanying narcolepsy (DSM-IV) (American Psychiatric Association, 1994), but still might elicit anxiety and distress; (iii) form of address of voices (talking in the first, second or third person), as particularly third person voices (talking about the patient and/or commenting on patient's activities, two of Schneider's first rank symptoms (Schneider, 1957)) are related to schizophrenic disorders.

Content of the AVHRS

The AVHRS is a structured 16-item interview, to evaluate patient's experiences during a predetermined period (e.g. past month, past year). Each item consists of a compulsory question, followed by optional support questions. Items are scored on a 5-point scale, ordered in increasing severity, facilitating calculation of a severity score. For experienced therapists and researchers, no training in administering the AVHRS is required. Duration of the interview is around 20 min.

Subjects

Data from two samples were analysed: (1) adult patients ($n=62$), who were questioned about their voices in the past month as part of the regular intake at the VOPD. Forty-two patients (68%) had a diagnosis in the schizophrenic spectrum, 11 (18%) had a mood disorder with psychotic features, nine patients had various diagnoses (e.g. borderline personality disorder, dissociative disorder); (2) non-clinical children with auditory hallucinations ($n=347$) from a population-based survey of 3870 children, aged 7 and 8 years (Bartels-Velthuis *et al.* 2010). AVHRS items and scores of the two samples are presented in Table 1. Characteristics of participants are presented in Table 2.

* Address for correspondence: Agna Bartels, PhD, University Medical Center Groningen, University Center for Psychiatry, PO Box 30.001 (CC72), 9700 RB Groningen, The Netherlands.
(Email: a.a.bartels@umcg.nl)

Table 1. The AVHRS: items and scores of different samples in percentages

Items and item categories	Adult patients (<i>n</i> = 62)	7-/8-year-old children (<i>n</i> = 347)
Assessment period	Past month	Past year
1a Number of voices mean (s.d.; range)	5.9 (6.4;1–30)	1.7 (1.6;1–10)
	% ^a	% ^a
1b Separately or simultaneously		
Always one voice	16	70
Several voices, speaking separately	37	18
Several voices, (occasionally) speaking simultaneously	47	12
2 Hypnagogic and/or hypnopompic voices^b		
Only voices when falling asleep	0	
Only voices during waking up	0	
Voices when falling asleep and during waking up	0	
Voices occur at all times	100	
3 Frequency		
Once a week or less	6	70
At least once a day	32	27
At least once an hour	16	1
Continuously	45	2
4 Duration		
A few seconds	7	66
A few minutes	29	29
At least one hour	8	3
Continuously	57	2
5 Localization		
Inside the head	44	55
Inside and outside the head	20	21
Outside the head, near the ears	16	6
Outside the head, further away	20	17
6 Loudness		
Whispering	21	66
Normal voice	34	25
Louder than normal voice	7	7
Shouting or screaming	39	3
7 Attribution of origin		
Internal only	28	50
Mainly internal	30	7
Mainly external	25	5
External only	17	12
No idea	0	27
8 Negative content		
Never unpleasant	3	44
Only occasionally unpleasant	7	21
More than occasionally but less than 50% unpleasant	15	14
≥50% unpleasant	37	12
Always unpleasant	39	10
9 Severity of negative content		
Not unpleasant	3	53
Unpleasant, but not about patient/family	8	14
Unpleasant, about behaviour of patient/family	7	12
Unpleasant, about patient/family themselves	24	16
Threatening commands or orders	58	5
10 Frequency of distress or suffering		
Never	5	58
Sometimes	19	22
Half of the time	15	5

Continued

Table 1. Continued

Items and item categories	Adult patients (<i>n</i> = 62)	7-/8-year-old children (<i>n</i> = 347)
Assessment period	Past month	Past year
Most of the time	37	11
Always	24	4
11 Intensity of distress or suffering		
None	7	63
Some	16	18
Serious	19	11
Severe	37	7
Extreme	21	1
12 Interference with daily functioning		
None	3	71
Limited	29	27
Moderate	26	1
Severe	23	1
Complete interference (hospitalization)	19	1
13 Control over voices		
Full control	3	16
Some control most of the time	27	12
Some control half of the time	10	3
Some control occasionally	37	9
No control	23	61
14 Anxiety		
Never	16	53
Hardly ever	18	13
Sometimes	20	27
Most of the time	10	6
At times completely in panic	36	1
15 Interference with thinking		
Never	15	44
Hardly ever	11	12
Sometimes	36	25
Most of the time	32	12
Always	7	7
16 Form of address^c		
16.1: first person voices present	30	31
16.2: second person voices present	92	66
16.3: third person voices present	53	9
Severity score mean (S.D.; range)	7.7 (3.6; 0–13)	2.1 (1.9; 0–10)

^aDue to rounding, percentages may not add up to 100.

^bThis item was discarded in the children study.

^cFirst, second and third person voices are not mutually exclusive.

Analysis

To establish inter-rater agreement, 23 successive interviews from sample 1 were observed by four raters. In sample 2, the Dutch Central Committee on Research involving Human Subjects did not allow inter-rater agreement assessments. Thus, each of the eight interviewers of sample 2 rated five DVD-recorded interviews of VOPD patients. Inter-rater agreement was analysed with Agree-6 for Windows (Popping, 1983).

Measure of agreement was (Cohen's) kappa (ranging from '0' = 'no agreement' to '1' = 'perfect agreement'). Linear weighted kappas were used, as item responses are ordered in increasing severity, except for the items on HHH and localization of voices: the main purpose of the item on HHH is to differentiate between HHH and more genuine AVH and the item 'localization of AVH' (hearing voices inside or outside the head) was shown to have no intrinsic effects on severity (Copolov *et al.* 2004).

Table 2. Characteristics of participants and psychometric outcomes of the AVHRS

Characteristics	Sample 1		Sample 2		Five-year follow-up study of Sample 2	
	Patients VOPD	Subsample of VOPD patients for inter-rater agreement	AVH+ children prevalence study		Subsample of VOPD patients for inter-rater agreement	
Number	62	23	347		5	
% Female	55	83	49		20	
	Mean	s.d. (Range)	Mean	s.d. (Range)	Mean	s.d. (Range)
Age (years)	37.2	12.6 (13–62)	38.7	11.4 (13–62)	8.0	0.46 (7–9)
Duration of AVH (years)	11.9	9.6 (0.5–36.0)	13.3	10.1 (0.5–30.0)	Not assessed	6.6
						5.9 (2.0–13.0)
Outcome measures	Internal consistency ^a (Cronbach's alpha)	Inter-rater agreement ^a (Cohen's kappa)	Internal consistency ^b (Cronbach's alpha)		Inter-rater agreement ^a (Cohen's kappa)	
	0.84	0.84 (4 raters)	0.77		0.88 (8 raters)	

VOPD, voices outpatient department; AVH+, with auditory vocal hallucinations; s.d., standard deviation.

^aPast month assessment.

^bPast year assessment.

Internal consistency was analysed using data of all available AVHRS interviews within the VOPD ($n = 62$) and within the sample of 7- and 8-year-old children ($n = 347$). Measure of internal consistency used was Cronbach's alpha.

To examine face validity, all participating adult patients were questioned about the comprehensibility and comprehensiveness of the scale in a systematic way using a short questionnaire. They were asked whether all questions were clear and understandable, and whether they missed any topic.

In order to test the discriminative power of the AVHRS a severity score was composed, by counting the two most severe scores of each item, except for those on HHH and on localization of voices.¹

Results

The inter-rater agreement scores (Cohen's kappas) were 0.84 and 0.88, respectively, internal consistency scores (Cronbach's alpha) were 0.84 and 0.77 (Table 2).

All patients very well understood the meaning of the questions, clearly recognized the content as part of their voice hearing and they did not mention any missing topic.

Mean severity score in the patient's sample was 7.7 (s.d. 3.6; range 0–13), with women scoring slightly higher than men (8.1 *v.* 7.3, not significant). Patients with a diagnosis in the schizophrenic spectrum had a

significantly lower AVH severity score (mean 6.8, s.d. 3.4) than the group with other diagnoses (mean 9.7, s.d. 3.3) ($t = 3.20$, $p = 0.002$).

Mean severity score in the children's sample was 2.1 (s.d. 1.9; range 0–10), with almost equal scores of girls and boys.

A significant difference between severity scores of patients' and children's samples was found ($t = 11.9$, $p = 0.000$).

Discussion

The AVHRS showed some good psychometric properties: the inter-rater agreement can be regarded as 'excellent' (Altman, 1995) and the internal consistency as 'good' (Feinstein, 1987).

Patients had much higher severity scores than children, indicating the discriminative validity of the instrument (children being non-clinical general population participants). Speculating on the difference in AVH severity score between schizophrenic spectrum and other patients, we observed that non-schizophrenic spectrum patients were referred to the VOPD only when previous treatment had little or no effect. Therefore, the group of non-schizophrenic spectrum patients might be regarded as a selective group with more enduring severe auditory hallucinations.

In daily clinical practice, the AVHRS has proven to be a useful instrument as a starting point for therapeutic interventions, as patients indicated that they were questioned about their voices in a clear, thorough

¹ For discriminative purposes, severity indexing in the child studies (Bartels-Velthuis et al. 2011, 2012) was less rigorous.

and recognizable way, thus adhering to therapy compliance (Cohen & Berk, 1985). Likewise, the AVHRS was used in recommended (early) family intervention programs (Onwumere *et al.* 2011) in which it was shown to be an effective outcome measure (Jenner & Van de Willige 2001; Van de Willige *et al.* 1996).

Some limitations are apparent. First, test-retest reliability of the AVHRS has not been examined. Test-retest analyses in the adult sample could not be performed, because in the clinical setting of the VOPD, the AVHRS assessment at intake is instantly incorporated into treatment. Thus, patients are prompted from the first contact to act upon their voices and to employ coping strategies to control them. When patients are reflecting on their voices, as a result, retest data will be contaminated with therapy instructions. Test-retest analyses in the children's sample could not be performed due to regulations of the Central Committee on Research involving Human Subjects in the Netherlands in order to avoid stigmatization. A second limitation is that, though the AVHRS has proven to measure effects of therapy effectively, sensitivity to change has not yet been assessed. Likewise, all 7-and-8-year-olds were assessed with the AVHRS. Third, generalizability of the results may perhaps be debatable. However, as the instrument showed to be suitable in two very divergent samples (chronic patients and children from the general population), it might be assumed that the AVHRS is suitable in a broad range of samples. The item scores presented in Table 1 illustrate the discriminative ability. Fourth, external validity of the scale has not yet been established. This may be a challenge for future research. Finally, a supposedly limitation of not performing factor analyses (like Haddock *et al.* 1999), would not add to better reliability of the scale, as in our opinion all topics of the AVHRS are important, making the outcome of the interview not reducible to a few major factors.

Strength of the study is that in the research period no 'less suitable' patients were excluded. Also, in the 5-year follow-up of the children study, the value of the severity index has proven to be useful: it was predictive for persistence of hearing voices (Bartels-Velthuis *et al.* 2011), higher level of severity was associated with stronger exposure to social adversity and level of severity also showed progressively stronger associations with delusions (Bartels-Velthuis *et al.* 2012).

Comparing the AVHRS with other instruments in this field, we consider that the MUPS (Carter *et al.* 1995) is a rather extensive instrument (356 items) for which empirically derived scales have not yet been developed. The revised Beliefs about Voices Questionnaire (BAVQ-R) (Chadwick *et al.* 2000) merely

provides insight into a patient's relationship with auditory hallucinations, and does not assess form or content of the voices. We suppose that, compared with the AHS (Haddock *et al.* 1999), the AVHRS with more and elaborate items might improve diagnostics and treatment of hearing voices. For instance, in clinical practice, it is useful to know whether a patient is hearing several voices, requiring different coping strategies per voice. Also, hearing more than one voice was shown to be a predictor of persistence of AVH in the 5-year follow-up study in children (Bartels-Velthuis *et al.* 2011). As voices may vary in degree of anxiety elicited, it might be a straightforward treatment policy to deal with least frightening voice(s) first. Furthermore, it is important to distinguish form of address, as first person voices might be regarded as patients' own thoughts, often related to trauma or negative self-image, which also require a specific therapeutic approach.

Considering the above and the conclusion of Ratcliff *et al.* (2011) that the PSYRATS (Haddock *et al.* 1999) might be refined and validated further, we consider that the AVHRS meets these recommendations and will be useful in identifying treatment targets beyond diagnosis (Nasrallah *et al.* 2011).

Copies of the AVHRS are available in English, Spanish and Dutch and can be downloaded from <http://www.rgoc.nl/#home/downloads>.

Acknowledgements

Participation of the patients and children in the study is gratefully acknowledged.

Conflict of Interests

None.

A. A. Bartels-Velthuis^{1,*}, G. van de Willige¹,
J. A. Jenner² and D. Wiersma¹

¹University of Groningen, University Medical Center Groningen, University Center for Psychiatry, Groningen, The Netherlands

²Jenner Consultants, Haren, The Netherlands

References

- Altman DG (1995). *Practical Statistics for Medical Research*. Chapman and Hall: London.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*. American Psychiatric Association: Washington, DC.

- Bartels-Velthuis AA, Jenner JA, Van de Willige G, Van Os J, Wiersma D** (2010). Prevalence and correlates of auditory vocal hallucinations in middle childhood. *British Journal of Psychiatry* **196**, 41–46.
- Bartels-Velthuis AA, Van de Willige G, Jenner JA, van Os J, Wiersma D** (2011). Course of auditory vocal hallucinations in childhood: 5-year follow-up study. *British Journal of Psychiatry* **199**, 296–302.
- Bartels-Velthuis AA, Van de Willige G, Jenner JA, Wiersma D, Van Os J** (2012). Auditory hallucinations in childhood: associations with adversity and delusional ideation. *Psychological Medicine*, **42**, 583–593.
- Carter DM, Mackinnon A, Howard S, Zeegers T, Copolov DL** (1995). The development and reliability of the Mental-Health Research-Institute Unusual Perceptions Schedule (MUPS) – an instrument to record auditory hallucinatory experience. *Schizophrenia Research* **16**, 157–165.
- Chadwick P, Lees S, Birchwood M** (2000). The revised beliefs about voices questionnaire (BAVQ-R). *British Journal of Psychiatry* **177**, 229–232.
- Cohen CI, Berk LA** (1985). Personal coping styles of schizophrenic outpatients. *Hospital and Community Psychiatry* **36**, 407–410.
- Copolov D, Trauer T, Mackinnon A** (2004). On the non-significance of internal versus external auditory hallucinations. *Schizophrenia Research* **69**, 1–6.
- Feinstein AR** (1987). *Clinimetrics*. Yale University Press: New Haven and London.
- Haddock G, McCarron J, Tarrier N, Faragher EB** (1999). Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scale (PSYRATS). *Psychological Medicine* **29**, 879–889.
- Jenner JA, Van de Willige G** (2001). HIT, hallucination focused integrative treatment as early intervention in psychotic adolescents with auditory hallucinations: a pilot study. *Acta Psychiatrica Scandinavica* **103**, 148–152.
- Jenner JA, Van de Willige G** (2002). *The Auditory Vocal Hallucination Rating Scale (AVHRS)*. University of Groningen, University Medical Center Groningen, University Center for Psychiatry: Groningen.
- Nasrallah H, Tandon R, Keshavan M** (2011). Beyond the facts in schizophrenia: closing the gaps in diagnosis, pathophysiology, and treatment. *Epidemiology and Psychiatric Sciences* **20**, 317–327.
- Onwumere J, Bebbington P, Kuipers E** (2011). Family interventions in early psychosis: specificity and effectiveness. *Epidemiology and Psychiatric Sciences* **20**, 113–119.
- Popping R** (1983). *Measures of Agreement for Nominal Data (in Dutch)*. PhD Thesis, University of Groningen: Groningen.
- Ratcliff K, Farhall J, Shawyer F** (2011). Auditory hallucinations: a review of assessment tools. *Clinical Psychology and Psychotherapy* **18**, 524–534.
- Schneider K** (1957). Primary & secondary symptoms in schizophrenia (in German). *Fortschritte der Neurologie-Psychiatrie* **25**, 487–490.
- Van de Willige G, Jenner JA, Wiersma D** (1996). *Effectiveness of HIT for Patients with Persistent Auditory Hallucinations [in Dutch]*. University of Groningen: Groningen.
- Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N** (1990). SCAN – Schedules for Clinical-Assessment in Neuropsychiatry. *Archives of General Psychiatry* **47**, 589–593.