A Comparison of Recordings of Sentences and Spontaneous Speech: Perceptual and Acoustic Measures in Preschool Children's Voices.

Anita McAllister and Signe Kofoed Brandt

Linköping University Post Print

N.B.: When citing this work, cite the original article.

A Comparison of Recordings of Sentences and Spontaneous Speech: Perceptual and Acoustic Measures in Preschool Children's Voices

*Anita McAllister and †Signe Kofoed Brandt, *Linköping and †Katrineholm, Sweden

Summary: A well-controlled recording in a studio is fundamental in most voice rehabilitation. However, this laboratory like recording method has been questioned because voice use in a natural environment may be quite different. In children's natural environment, high background noise levels are common and are an important factor contributing to voice problems. The primary noise source in day-care centers is the children themselves. The aim of the present study was to compare perceptual evaluations of voice quality and acoustic measures from a controlled recording with recordings of spontaneous speech in children's natural environment in a day-care setting. Eleven 5-year-old children were recorded three times during a day at the day care. The controlled speech material consisted of repeated sentences. Matching sentences were selected from the spontaneous speech. All sentences were repeated three times. Recordings were randomized and analyzed acoustically and perceptually. Statistic analyses showed that fundamental frequency was significantly higher in spontaneous speech ($P < 0.01$) as was hyperfunction ($P < 0.001$). The only characteristic the controlled sentences shared with spontaneous speech was degree of hoarseness (Spearman's rho = 0.564). When data for boys and girls were analyzed separately, a correlation was found for the parameter breathiness (rho = 0.551) for boys, and for girls the correlation for hoarseness remained (rho = 0.752). Regarding acoustic data, none of the measures correlated across recording conditions for the whole group.

Key Words: Recording conditions–Spontaneous speech–Children–Perceptual ratings–Voice quality–Acoustic measures.

INTRODUCTION

In a clinical setting, problems related to voice function are routinely assessed by perceptual evaluations of voice quality.[1] The most common material for this assessment is a standardized recording, consisting of reading a text aloud, naming pictures or repeating sentences, and sustaining vowels depending on the patient population. The recordings are often carried out in a sound-treated booth aiming at high-quality recordings. Based on these recordings, a perceptual assessment of voice quality along different perceptual parameters is carried out.[2–4] The result of the perceptual evaluation together with laryngeal status makes up the basis for decisions regarding intervention. Also, improvement in voice quality

is one of the primary benchmarks against which treatment outcome is evaluated commonly, assessed by an evaluation along perceptual and acoustic parameters after completed intervention.

A dysfunctional voice may be a serious social and psychological problem for adults[5] and children.[6,7] Many habits, including vocal habits, are probably established during childhood. Thus, undesirable vocal habits may originate during early childhood and continue into adult life.[8,9] This would point to the importance of voice research focusing on child's voice and the treatment and prevention of voice disorders in children. This research also needs to include vocal behavior and vocal demands in children's everyday life.

Recently, the importance of in situ recordings of natural vocal behavior in everyday life situations has been pointed out.[10–13] In a study of preschool teachers' voices, Södersten et al[14] compared mean fundamental frequency (F0) in a controlled recording to the F0 in the spontaneous speech during work. They found that the mean F0 was higher in the work related recording compared with the controlled condition, indicating that a controlled recording may not reflect mean F0 in spontaneous speech under natural conditions. There are few studies of children's voice use in a natural setting. In a recent study of mean F0 in children's and teachers' voices in a preschool setting, the results showed a significant difference for both children and adults between the recordings of sentences compared with real work/play situations.[15] The findings support the conclusion that controlled setups are not suitable to evaluate F0 values in a natural setting. These findings have also been supported by two studies of preschool-aged children, a case study of a 5-year-old boy[16] and a study comparing F0 in children at play compared with structured situations.[17] The results indicate that studio recordings in a clinical setting need to be complemented by recordings in real-life situations and environments to correctly assess habitual F0 in children.

The question asked in the present study is, does this difference in vocal behavior regarding F0 also apply to other aspects of voice? Thus, children's voice quality, mean F0, and perturbation in a controlled recording were compared with sentences obtained during regular activities at the day-care center (DCC).

METHODS
Subjects
Recordings from eleven 5-year-old children in a previous study on environmental factors contributing to voice problems in children were selected.[18] The children had no history of hearing or speech problems, or frequent ear, nose, and throat infections. No initial survey of voice quality was made. The children attended three DCCs in a city with approx. 135 000 inhabitants at the time of the data collection, situated 200 km south of Stockholm in Sweden. An informed consent form was signed by the parents before the recording of the children. If the children themselves declined participation, no recording was made.

Recording method

A binaural recording technique was used to be able to separate the subject's voice from surrounding background noise.[19] A DAT recorder and two omnidirectional electret condenser microphones (TCM 110), AA-video, Linköping, Sweden were used. The microphones were taped to the cheek directly in front of the ears on each child at equal distance from the mouth. The mouth-to-microphone distance varied between 4 and 6 cm. All microphone distances were normalized to 30 cm before analyses. The children were recorded three times during a normal day at the DCC, on arrival, during lunch, and during play in the afternoon. All recordings were gathered by the same test leader. Each recording session started with a calibration of the loudness level. Because of microphone problems, two recordings from a girl had to be discarded leaving a total of 31 recordings.

The children's voices were separated from the background noise using the software Aura (a custom-made program by Svante Granqvist).[19] The resulting data were then separated into several files consisting of background noise, controlled speech, and spontaneous speech, respectively. The mean background noise levels varied between 81.1 and 85.4 dBA.

Speech material

The controlled speech material composed of three short sentences consisting of sonorants only (En blå bil. En gul bil. En röd bil). To reduce effects of imitation of intonation or pitch level, the instruction was ''Can you say a blue car? A yellow car. A red car.'' When the child had done this once, he or she was asked to say it again three times. The last two repetitions were used in the analysis. Each recording started with the repetitions of these short sentences recorded in a silent room at the preschool, the recording then continued with spontaneous speech during their regular daily activities for 45–60 minutes. No instructions were provided regarding specific activities or vocal behavior. All recordings were made indoors.

To compare voice quality in the controlled sentences, short sentences were also selected from each child's separated recordings of spontaneous speech production from the same recording session. The selected spontaneous samples were chosen to be as neutral in loudness and F0 as possible. Thus, sections with shouting or elevated F0 were disregarded to avoid clear differences in speech conditions. No consideration regarding content of the sample was made. Thus a total of 62 samples, 31 of controlled sentences and 31 of spontaneous speech were obtained. The duration of the compared samples was similar.

Perceptual evaluation

The randomized recordings from the repeated sentences and the selected samples from the spontaneous speech for each child were perceptually analyzed by a group of three expert listeners on two separate occasions separated in time by more than 1 year. All three

listeners were speech and language pathologists with extensive experience in working with voice disorders for a minimum of 15 years. In the evaluation protocol, all parameters were represented by a 100-mm visual analog scale (VAS) with ''not at all'' and ''a lot'' at the extremes, except for the parameter pitch, represented by a 200-mm line with ''very high'' and ''very low'' at the extremes.[18] The other parameters were hoarseness, breathiness, roughness, hyperfunction, and an open parameter to offer options for the raters, if necessary. Three repetitions of each set of sentences for each condition, controlled sentences or sentences selected from spontaneous speech were presented to the listener on a computer using Sennheiser PX200 headphones. The instruction to the listeners was to assess the three repeated sentences as one sample.

Acoustic measures
 Acoustic measures F0 and perturbation percent were measured using the software Soundswell signal workstation 4.5 (Saven Hightech).[20] The obtained speech signal was filtered using a high pass filter set at 50 Hz and a low pass filter set at 550 Hz. Maximum F0 was adapted to the highest F0 frequency occurring in the speech sample. Mean and standard deviation (SD) of F0 were calculated from the F0 trend. Also, mean and SD for the perturbation measures were calculated using the software Soundswell.[20]

Statistical analysis
Statistical analyses were performed using SPSS_Windows version 18.0 (IBM. SPSS Inc, Chicago, IL) and a two-tailed Pearson's or Spearman's rank correlation depending on parametric or nonparametric data. The significance level was set at $P \leq 0.05$. Significance for perceptual data was evaluated using Wilcoxon signed-rank test and for acoustic measures a twotailed, pair-wise Student's t test.

Ethical approval
Before data collection, ethical approval was received from the Regional Ethical Research Committee at Linköping University, no. 03-173.

RESULTS
Interrater agreement for the perceptual evaluation was calculated using a Spearman's rho correlation. The agreement between judges was satisfactory. For the controlled sentences, the agreement varied between rho = 0.81 and 0.89 for the different parameters with the highest agreement for the parameter hyperfunction. For the spontaneous speech sentences, the agreement was somewhat higher varying between rho = 0.90 and 0.95 with the highest value for the parameter hoarseness. The perceptual evaluation showed overall higher values in the sentences selected from spontaneous speech compared with the controlled conditions for all perceptual parameters (Figure 1). This difference was significant for the parameter hyperfunction at $P < 0.001$ according to a Wilcoxon signed rank test.

Mean F0 for all children was somewhat higher in spontaneous speech compared with the sentences, 306 and 282 Hz, respectively. This difference was significant according to the two-tailed Student's t test (Figure 2).

The mean perceptual ratings of the repeated sentences and the selected sentences from spontaneous speech for all children showed a correlation across speech conditions only for the parameters hoarseness according Spearman's rho (Table 1). When separating the recordings of boys and girls, the correlation across speech conditions for hoarseness remained for girls but for boys the parameter breathiness reached significance (Table 2).

Regarding the acoustic measures, mean F0 and perturbation percent, no correlation was found across speech conditions for the whole group. When analyzing data for boys and girls separately, still no correlation was found for F0, but for perturbation a correlation was found for both boys and girls according to Pearson's correlation. For girls, the correlation was dependent on one value. When this outlier was removed, the correlation disappeared (Table 3).

DISCUSSION
In the present study, the relationship between acoustic measures and a perceptual evaluation of controlled recordings of repeated sentences, and sentences selected from spontaneous speech were investigated. The data were obtained from recordings of 11 children on the same day and in the same environment. Selected samples were chosen to be as similar as possible. Thus, sections with shouting or obviously elevated F0 were disregarded to avoid clear differences in the compared samples. Comparisons included F0, perturbation in percent, and also perceptual measures of voice quality. Previous studies have mainly focused on comparing changes in F0 in a controlled recording compared with spontaneous speech.[14–17]

Results showed that when analyzing all children, there was a correlation between perceptual voice quality in a controlled recording compared with spontaneous speech for the parameter hoarseness. Hoarseness has previously been found to be a stable concept in children's voice quality both for professional raters and laymen.[21,22] This study corroborates these findings. However, when analyzing the group according to gender a slightly different pattern emerged. For boys, the controlled sentences were representative of vocal behavior in spontaneous speech for the parameter breathiness only. For girls, the correlation for hoarseness remained. This seems to indicate that a breathy voice is a more constant vocal characteristic in boys' voices, a feature not influenced by changes in F0 or vocal loudness. In adult's voices, breathiness is regarded as a key characteristic of the female voice.[23,24] However, in children's voices it is a common characteristic in both boys' and girls' voices.[22] It is possible that the present finding could be related to vocal fatigue or the presence of a slight vocal fold edema in some of the recorded boys. Regrettably, these speculations cannot be confirmed because laryngeal inspection was not carried out at the time of the recording.

Regarding acoustic measures, the findings showed that mean F0 was lower in the controlled sentences compared with the spontaneous speech. This is in accordance with previous studies on F0, indicating that a controlled setup is unfit to evaluate F0 values in real-life situations.[14–17] Higher F0 values in spontaneous speech are likely because of an increase in vocal loudness [24,25] that in turn may be related to increased background noise.[18,26] The preschool environment is often noisy.[14,18] This was also true for the three preschools in the present study varying between mean values of 81.1 and 85.4 dBA, both values above 80 dBA, where noise reductive actions are required according to Swedish work environmental regulations (AFS 2006). However, children and their activities are the primary noise source in preschools. In a previous study on noise and voice characteristics in a preschool setting, a link was found between high background noise levels and higher ratings on perceptual voice quality measures in children.[18]

For the perturbation measure, a correlation was found between the two speech conditions for the boys. For girls, the correlation was dependent on one outlier value. When this value was removed, the correlation disappeared. However, both perturbation values for this girl were higher than for the other children indicating a somewhat dysfunctional voice. The perceptual evaluation showed fairly high ratings of hoarseness.

Perceptual and acoustic evaluations based on voice recordings in a natural setting are often limited in the analysis because of background noise. This has lead to the development of different portable voice accumulators.[11,27–29] The aim of portable voice collectors is to obtain representative samples of voice use in a natural setting and holds the potential to increase our knowledge of natural vocal behavior in both patients and the vocally healthy. For patients, these recordings can provide clinicians with important additional information on conditions and demands influencing spontaneous vocal behavior and vocal loading.

CONCLUSION

The evaluation of voice quality, F0, and perturbation in standard sentences and sentences selected from spontaneous speech was compared. A total of 62 samples from 11 children were analyzed. Data showed a correlation between the standard sentences and sentences selected from spontaneous speech for the voice quality parameter hoarseness only. F0 was significantly higher in spontaneous speech. For boys, there was a correlation across speech tasks for the parameters breathiness and perturbation (%) and for girls for hoarseness. The findings suggest that controlled recording conditions may be unsuitable to approximate children's vocal behavior in a natural setting.

REFERENCES

1. DeBodt MS, Wuyts FL, Van de Heyning PH, Croux C. Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. J Voice. 1997;11:74–80.

2. Isshiki N, Okamura H, Tanabe M, Morimoto M. Differential diagnosis of hoarseness. Folia Phoniatr (Basel). 1969;21:9–19.

3. Hammarberg B. Perceptual and acoustic analysis of dysphonia [PhDthesis]. Stockholm, Sweden: Karolinska Institute, Department of logopedics
and phoniatrics; 1986.

4. Hammarberg B. Voice research and clinical needs. Folia Phoniatr Logop. 2000;52:93–192.

5. Benninger MS, Ahuja AS, Gardner G, Grywalski C. Assessing outcomes for dysphonic patients. J Voice. 1998;12:540–550.

6. Nienkerke-Springer A, McAllister A, Sundberg J. Effects of family therapy on children's voices. J Voice. 2005;19:103–113.

7. Zur KB, Cotton S, Kelchner L, Baker S,Weinrich B, Lee L. Pediatric Voice Handicap Index (pVHI): a new tool for evaluating pediatric dysphonia. Int J Pediatr Otorhinolaryngol. 2007;71:77–82.

8. Zerffi WAC. Functional vocal disabilities. Laryngoscope. 1939;49: 1143–1147.

9. Powell M, Filter MD,Williams B. A longitudinal study of the prevalence of voice disorders in children from a rural school division. J Commun Disord. 1989;22:375–382.

10. Szabo A, Hammarberg B, H_akansson A, S€odersten M. A voice accumulator device: evaluation based on studio and field recordings. Logoped Phoniatr Vocol. 2001;26:102–117.

11. Szabo A, Hammarberg B, Granqvist S, S€odersten M. Methods to study preschool teachers' voice at work: simultaneous recordings with a voice accumulator and a DAT recorder. Logoped Phoniatr Vocol. 2003;28:29–39.

12. Ternström S, Södersten M, Boman M. Cancellation of simulated environmental noise as a tool for measuring vocal performance during noise exposure. J Voice. 2002;16:195–206.

13. Vilkman E. Occupational safety and health aspects of voice and speech professions. Folia Phoniatr Logop. 2004;6:220–253.

14. Södersten M, Granqvist S, Hammarberg B, Szabo A. Vocal behaviour and vocal loading factors for pre-school teacher at work studied with binaural DAT recordings. J Voice. 2002;3:356–371.

15. Lindstrom F, Ohlsson AC, Sj€oholm J, Waye KP. Mean F0 values obtained through standard phrase pronunciation compared with values obtained from the normal work environment: a study on teacher and child voices performed in a preschool environment. J Voice. 2010;24:319–323.

16. Hunter EJ. A comparison of a child's fundamental frequencies in structured elicited vocalizations versus unstructured natural vocalizations: a case study. Int J Pediatr Otorhinolaryngol. 2009;73:561–571.

17. Chen Y, Kimelman MDZ, Micco K. Investigation of habitual pitch during freeplay activities for preschool-aged children. Int J Pediatr Otorhinolaryngol. 2009;73:73–80.

18. McAllister A, Granqvist S, Sjölander P, Sundberg J. Child voice and noise: a pilot study of the effect of a day at the day-care on ten children's voice quality according to perceptual evaluation. J Voice. 2009;23:587–593.

19. Granqvist S. The self-to-other ratio applied as a phonation detector for voice accumulation. Logoped Phoniatr Vocol. 2003;28:71–80.

20. Soundswell Signal Workstation_, 4.5. Sweden: Saven Hitech; 2007.

21. Sederholm E, McAllister A, Sundberg J, Dalkvist J. Perceptual evaluation of hoarseness using continuous scales. Scand J Logop Phoniatr. 1993;18: 73–82.

22. McAllister A, Sederholm E, Sundberg J, Gramming P. Relations between voice range profiles and physiological and perceptual voice characteristics in ten-year-old children. J Voice. 1994;3:230–239.

23. Biever D, Bless D. Vibratory characteristics of the vocal folds in young adult and geriatric women. J Voice. 1989;3:120–131.

24. Södersten M, Lindestad P. Glottal closure and perceived breathiness during phonation in normally speaking subjects. J Speech Hear Res. 1990;33: 601–611.

25. Gramming P, Sundberg J, Ternström S, Leandersson R, PerkinsW. Relationship between changes in voice pitch and intensity. J Voice. 1998;2:118–126.

26. Södersten M, Ternström S, Bohman M. Loud speech in realistic environmental noise: phonetogram data, perceptual voice quality, subjective ratings, and gender differences in healthy speakers. J Voice. 2005;19:29–46.

27. Ohlsson AC, Brink O, L€ofqvist A. A voice accumulation—validation and application. J Speech Hear Res. 1989;32:451–457.

28. Cheyne HA, Hanson HM, Genereux RP, Stevens KN, Hillman RE. Development and testing of a portable vocal accumulator. J Speech Lang Hear Res. 2003;46:1457–1467.

29. Lindström F, Persson Waye K, Södersten M, McAllister A, Ternström S. Background noise in a pre-schools and average sound pressure level and fundamental frequency of teachers' voices. J Voice. 2011;25:166–172.
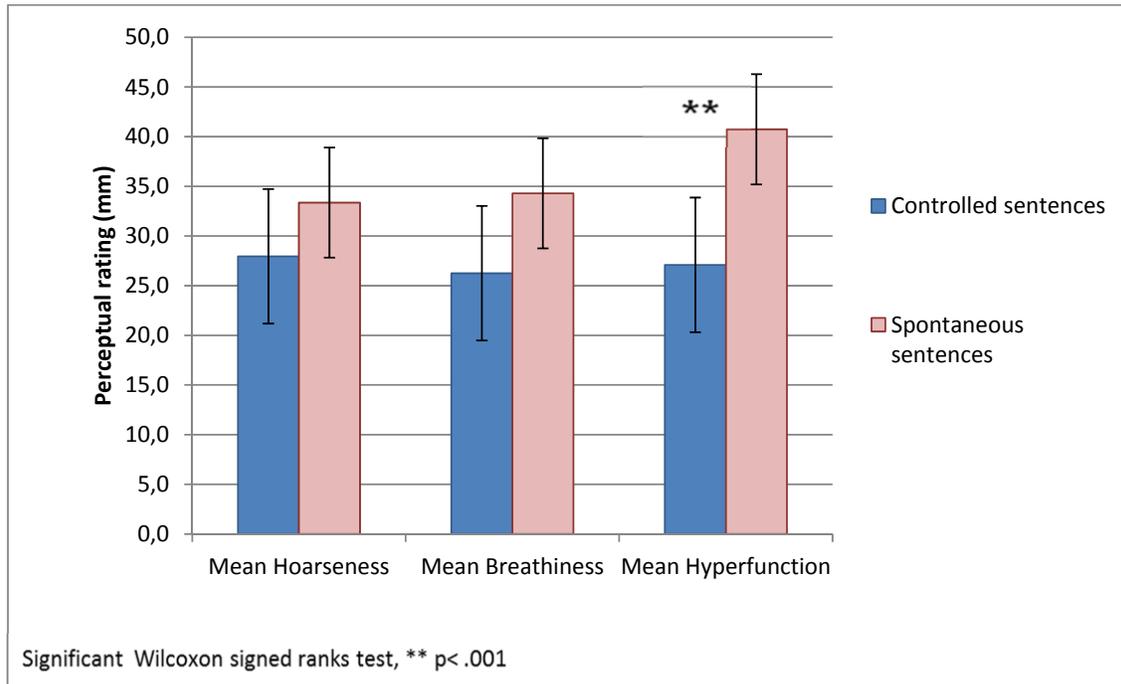
Figures and tables with legends

FIGURE 1. Mean perceptual ratings of voice quality along a 100-mm VAS for the parameters hoarseness, breathiness, and hyperfunction in all children based on controlled sentences and sentences selected from spontaneous speech. The difference between recording conditions was significant for the parameter hyperfunction.
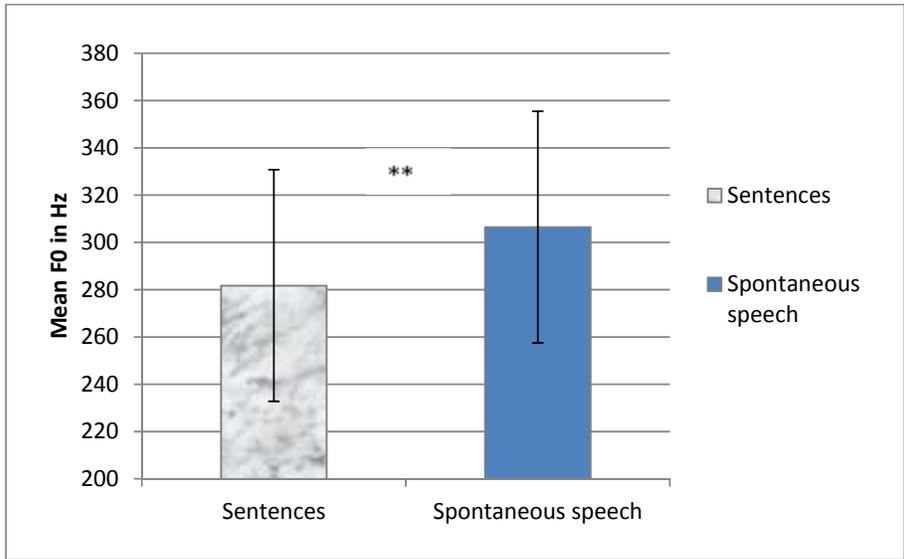
FIGURE 2. Mean F0 (Hz) for all children in controlled sentences and sentences selected from spontaneous speech. **P < 0.01.

Table 1. Correlations between the mean perceptual ratings of repeated sentences and spontaneous speech for all children according to a correlation using Spearman's rho.

		Hoarseness spont	Breathiness spont	Hyperfunction spont
Hoarseness sent	Correlation coeff	**0.564**[*]		
	Sig. (2-tailed)	**0.001**		
	N	**31**		
Breathiness sent	Correlation coeff		0.261	
	Sig. (2-tailed)		ns	
	N		31	
Hyperfunction sent	Correlation coeff			0.117
	Sig. (2-tailed)			ns
	N			31

**. Correlation is significant at the 0.01 level (2- tailed).

Table 2. Correlations between the mean perceptual ratings of repeated sentences and spontaneous speech for boys and girls respectively according to a correlation using Spearman's rho, n=number of analyzed recordings for each condition.

		Hoarseness sentences vs spontaneous speech	Breathiness sentences vs spontaneous speech	Hyperfunction sentences vs spontaneous speech
Boys	Correlation coeff	0.389	**0.551**[*]	-0.116
	Sig. (2-tailed)	ns	**0.018**	ns
	n	18	18	18
Girls	Correlation coeff	**0.752**[**]	0.426	0.412
	Sig. (2-tailed)	**0.003**	ns	ns
	n	13	13	13

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 3. Correlations between acoustic data from repeated sentences and spontaneous speech for boys and girls respectively according to a Pearson's correlation. n=number of analyzed recordings for each condition.

		F0 sentences vs spontaneous speech	Perturbation sentences vs spontaneous speech
Boys	Pearson Correlation	0.254	**0.727**[*]
	Sig. (2-tailed)	ns	**0.001**
	n	18	18
Girls	Pearson Correlation	0.327	-0.154
	Sig. (2-tailed)	ns	ns
	n	12	12

*. Correlation is significant at the 0.01 level (2-tailed).