A Preliminary Investigation of Potential Biases in Phonation Threshold Pressure Analysis

Anusha Sundarrajan, [1]Elizabeth Erickson-Levendoski, and M. Preeti Sivasankar, *West Lafayette, Indiana*

Summary: Objective. Phonation threshold pressure (PTP) is a voice measure used in both research and clinic. PTP data analysis is susceptible to bias from investigator awareness of experimental hypothesis, and poor investigator training. The objective of this study was to systematically examine the role of these two biases on PTP data analysis.

Study Design. Prospective design.

Methods. Two trained investigators analyzed PTP datasets. The datasets were identical, but uniquely labeled so that the investigators were not aware that the datasets contained the same data. Each investigator analyzed two datasets. For one dataset, investigators were "blinded" to the experimental hypothesis. For the other dataset, the investigators were "unblinded" and provided a fake experimental hypothesis. Intraclass correlations were used to examine intrarater and interrater reliability.

Results. For both investigators, intraclass correlations within the excellent range were obtained for intrarater reliability. In contrast, lower intraclass correlations were obtained for interrater reliability.

Conclusions. The high intrarater reliability obtained in this preliminary study suggests that awareness of experimental hypothesis may not significantly bias PTP analysis. Conversely, lower interrater reliability is indicative of differences between investigators analyzing the same data. Our findings contribute to the growing body of literature that seeks to standardize the use of PTP in research and the clinic. Future investigations are needed to identify methods to improve interrater reliability and that quantify the effects of biases on PTP data collection.

Key Words: Phonation threshold pressure–Biases–Reliability–Intraclass correlations.

INTRODUCTION

Phonation threshold pressure (PTP) is the minimum lung pressure required to initiate and sustain vocal fold oscillation.[1] PTP is commonly used in research laboratories to noninvasively assess changes in laryngeal biomechanics[2–6] but is less ubiquitous in the clinic. One obstacle in the widespread adoption of PTP[7] is susceptibility to bias. PTP data collection and analysis retain some subjectivity. Investigator knowledge of the hypothesis and experimental conditions and investigator training and experience are potential sources of bias in data collection and analysis. These biases may be manifested as inconsistent cuing for evoking threshold phonation, providing inadequate feedback about productions, and selective peak picking. Blinding, defined as the deliberate withholding of information from participants and/or investigators involved in a study[8,9] is used to reduce biases in PTP data collection and analysis.

Double-blinded research designs have been used in studies of PTP. Verdolini et al[2] documented the role of systemic and secretory dehydration in altering PTP with a double-blinded placebo controlled study. Tanner et al[3] used a double-blinded within crossover design, to examine the effects of nebulized treatments on PTP following a surface laryngeal dehydration challenge in classically trained sopranos. Erickson-Levendoski and Sivasankar[10] investigated the adverse phonatory effects of caffeine with a double-blinded sham controlled experiment. Double blinding however, may not be applicable in all PTP studies. Complete participant and investigator blinding may be precluded in research studies that investigate the effects of environmental background noise or different ambient humidities and temperatures on PTP.[11–13] In such circumstances, investigator blinding during data analysis alone may be used.[14] However, to fully understand the importance of investigator blinding in PTP data analysis, it is essential to determine whether awareness of the experimental hypothesis can bias PTP analysis. In addition, it is useful to quantify how much investigator training is needed to accurately measure PTP.

The overall goal of this preliminary study was to investigate potential biases in PTP analysis. The primary focus was to determine whether knowledge of the experimental hypothesis can influence PTP analysis. The secondary focus was to compare the accuracy of PTP analysis between two investigators, trained simultaneously by the same researcher. We hypothesized that knowledge of the experimental hypothesis would bias PTP analysis and furthermore, that PTP data analyzed by similarly trained investigators would be in close agreement. To test these hypotheses, two trained investigators analyzed PTP data that were collected *a priori* by the authors. The investigators analyzed coded data and were blinded or unblinded to a fake, experimental hypothesis. We propose that knowledge from this study will inform practices for training investigators to analyze PTP data in both research and clinical settings.

METHODS

The Purdue University Institutional Review Board approved all the procedures used in the study.

Accepted for publication July 3, 2014.

This research was presented at the Annual Symposium: Care of the Professional Voice, May 28—June 1, 2014, Philadelphia, Pennsylvania.

From the Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana.

[1]Currently at University of Wisconsin-Madison, Division of Otolaryngology-Head & Neck Surgery, Department of Surgery, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin.

Address correspondence and reprint requests to M. Preeti Sivasankar, Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN 47907. E-mail: msivasankar@purdue.edu

Journal of Voice, Vol. 29, No. 1, pp. 22-25

0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2014.07.001

Investigators

PTP data were analyzed by two investigators who did not participate in data collection. The investigators read several published articles on PTP and discussed them with the senior author (M.P.S.). PTP training was provided to the investigators by the senior author and the two investigators practiced PTP analysis.

Data categorization and instructions

Four PTP datasets were created by the senior author. Each dataset contained the same data, which included 606 /pi/ pressure peaks. The datasets were labeled blinded1 (BL1), blinded2 (BL2), unblinded1 (UB1), and unblinded2 (UB2). The first investigator analyzed UB1 and BL1. The second investigator analyzed UB2 and BL2. The following instructions were provided to the investigators. For the UB1 dataset, "This dataset contains PTP data for 6 subjects (subject 1 through subject 6). Each subject performed the PTP task before loud reading and after loud reading. So each subject has PTP-pre data and PTP-post data. We want to test the hypothesis that loud reading increases PTP. This has been demonstrated before in PTP studies. So we would expect that PTP is higher for PTP-post data as compared to PTP-pre data." For the BL1 dataset the instructions were as follows: "This dataset contains PTP data for 12 subjects (subject 13 through subject 24). Please analyze these PTP data." The second investigator was provided with the same instructions for their unblinded (UB2) and blinded (BL2) datasets except for UB2 they were informed that they were analyzing subjects 37–48, and for BL2 they were requested to analyze PTP data from subjects 25–36. We chose a loud reading task as the fake hypothesis for the unblinded dataset because this loading challenge is commonly incorporated in PTP studies and there is a clear direction of experimental change that the investigators were aware of.[4,15] Each investigator analyzed their assigned datasets and their analyses were compared for interrater reliability and intrarater reliability.

PTP instrumentation, data collection, and analysis

The instrumentation for PTP data collection included a circumferentially-vented pneumotachograph face mask coupled to transducers (Glottal Enterprises, Syracuse, NY) for the measurement of oral flows and PTP. PTP data were collected using validated methods from six individuals (2 males and four females) in a humidity (50 ± 10%) and temperature (70 ± 5°F) controlled environment (Traceable Memory Hygrometer, VWR, Radnor, PA). The subjects were trained on PTP using established methods.[16] In brief, subjects produced 7–10 repetitions of /pi/ as softly as possible in a single breath at 92 beats/minute measured using a Seiko Digital Metronome (Model# DM50; Seiko Sports Life Co., Ltd, China). These repetitions constituted one string. Subjects produced 8–10 strings at C4 frequency as cued on a keyboard. The subjects were provided visual feedback and investigator modeling. PTP data were deemed accurate when meeting the following criteria: /pi/ pressure peaks of equal height as assessed visually, oral flows $\leq$15 mL/s during the /p/ production. For PTP analysis, the investigators analyzed the pressures for three consecutive /p/ peaks from each string. The first and last pressure peaks were not considered for inclusion. These peak pressures were averaged to yield PTP.

Statistical analysis

PTP data were summarized as means ± Standard deviation (SD). Data were analyzed using *SPSS statistical software* (IBM Version 20, Chicago, IL). To obtain intrarater reliability a two-way mixed, absolute, single-measures intraclass correlation coefficient (ICC) was calculated. Separate ICCs were performed for each investigator. For investigator 1, intrarater reliability was the ICC between UB1 and BL1. For investigator 2, intrarater reliability was the ICC between UB2 and BL2. To obtain interrater reliability, a two-way mixed, absolute, single-measures ICC was calculated. Separate ICCs were obtained for the blinded and unblinded datasets. For the blinded dataset, interrater reliability was the ICC between BL1 and BL2. For the unblinded dataset, interrater reliability was the ICC between UB1 and UB2.

RESULTS

The primary purpose of this study was to investigate whether knowledge of the experimental hypothesis would affect PTP analysis. We compared the PTP analysis for investigators who analyzed identical PTP data that were labeled differently (ie, BL1 and UB1 for investigator 1, and BL2 and UB2 for investigator 2). PTP results are depicted in Table 1. Mean PTP data for investigator 1 were 3.96 cm H_2O for the BL1 condition and 3.86 cm H_2O for the UB1 condition, yielding a mean difference of 0.1 cm H_2O. Mean PTP data for investigator 2 were 3.62 cm H_2O for the BL2 condition and 3.66 cm H_2O for the UB2 condition, yielding a mean difference of 0.04 cm H_2O. We also computed intrarater ICC for each investigator. ICCs are classified as excellent in the range of 0.75–1. Intrarater ICC for investigator 1 was excellent at 0.92. Intrarater ICC for investigator 2 was excellent at 0.96.

The second purpose of this study was to investigate the extent of agreement between two investigators who were trained simultaneously in PTP analysis. We predicted that these investigators would be in close agreement in their analysis of identical data (Table 1). The mean PTP data were 3.96 cm H_2O and 3.62 cm H_2O for BL1 and BL2 datasets respectively, yielding a mean difference of 0.34 cm H_2O. Mean PTP data were 3.86 cm H_2O and 3.66 cm H_2O for UB1 and UB2 datasets

TABLE 1.
Interrater and Intrarater Intraclass Correlation Coefficients (ICC) for Blinded and Unblinded Datasets for Each Investigator

Datasets	Investigator 1 PTP (cm H_2O) Mean ± SD	Investigator 2 PTP (cm H_2O) Mean ± SD	Interrater ICC
Blinded	3.96 ± 0.55	3.62 ± 0.57	0.76
Unblinded	3.86 ± 0.62	3.66 ± 0.61	0.78
Intrarater ICC	0.92	0.96	