

The Effects of Caffeine on Voice: A Systematic Review

*Vasilis L. Georgalas, *Niki Kalantzi, †Isolde Harpur, and *Ciarán Kenny, *†*Dublin, Ireland*

Summary: Background. Caffeine is considered a dehydrating agent due to its diuretic effects and influences the body's fluid balance. The relationship between voice and hydration has been widely investigated and it is accepted that inadequate hydration has detrimental effects on phonation. Since dehydration negatively affects the vocal folds and caffeine is considered a dehydrating agent, it can be hypothesized that voice might be negatively affected by caffeine intake. This systematic review aims to summarize and appraise the available evidence regarding the effects of caffeine on voice.

Methods. Randomized and non-randomized experimental studies of healthy participants were retrieved following an electronic searching of six databases in June 2020. No publication, language or date restrictions were applied. Data extraction of relevant data and risk of bias assessment was conducted independently by two reviewers.

Results. Five non-randomized experimental studies were deemed eligible for inclusion. The format of the administered interventions in the included studies was either liquid (coffee) or solid (caffeine tablets). Reported outcome measures used to examine the effects of caffeine on phonation consisted of acoustic, aerodynamic and (auditory & self-) perceptual. No measures were adversely affected by caffeine consumption.

Conclusion. Clinicians commonly advise patients to refrain from caffeine, as caffeine intake increases diuresis with subsequent effects on fluid balance. Such imbalances can potentially induce dehydration which can be detrimental to phonation. This notion cannot be supported empirically, as the evidence is deemed unreliable and no firm conclusions can be elicited to guide clinical practice. The results of this review demonstrate the lack of research in the field and the necessity for future investigations in order to inform evidence-based practice through reliable and valid outcomes.

Key Words: Caffeine–Voice– Hydration– Voice quality–Phonation– Systematic review.

INTRODUCTION

Caffeine

Caffeine is one of the most consumed substances in the world¹. It consists of three central nervous system (CNS) stimulants; paraxanthine (84%), theobromine (12%) and theophylline (4%).^{2,3}

Caffeine consumption has been linked with a lower risk of particular types of cancer, minimized risk of Type II diabetes and a reduced risk of developing Parkinson's disease, Alzheimer's disease and depression.⁴ In spite of its beneficial effects, consumption has also been associated with bone loss, reduced bone density, increased pregnancy risks, behavioural changes and sleep deprivation.^{5,6} The impact of caffeine on the body's fluid balance as a result of increased diuresis has been investigated. A review by Maughan & Griffin⁷ reported that 300mg of caffeine can induce diuresis, while a more recent meta-analysis suggested that 300mg of caffeine intake does not influence diuresis.⁸ Three studies have demonstrated that higher dosages (>450 mg) of caffeine intake have subsequent effects on total body water (TBW) volume, fluid balance and urine output volume.⁹⁻¹¹

Conversely, two studies reported to hydration status alterations were not noticed following consumption of moderate dosage of caffeine (244 mg-370 mg).^{12,13}

Side-effects of caffeine include diuresis, increased alertness and sleep deprivation, individuals can develop tolerance to these. The degree of tolerance varies amongst individuals and depends on an individual age and sex¹⁴. According to a review by Nehlig,¹⁵ the metabolic, pharmacokinetic, functional and physiological effects of caffeine may vary due to age, sex, diet, lifestyle and genetic factors.

Caffeine consumption is noted to be increasing worldwide, since caffeine is contained in numerous sources such as coffee, tea, chocolate, sodas, energy drinks and medications.¹⁶ Despite a lack of consensus regarding safe levels of intake, Health Canada (HC) provide advice about the amount of caffeine that is considered safe to be consumed¹⁷ with values based on the review by Nawrot et al.¹⁸ (Table 1).

Hydration and voice

Hydration refers to the TBW concentrations in the human body. Hydration status is described with the following terms: euhydration, dehydration and hyperhydration. Euhydration is the presence of fluid equilibrium and refers to adequate hydration levels within the human body.¹⁹ Research has shown that age, sex, adiposity levels and population characteristics (eg, occupation, ethnicity) are some of the factors contributing to different TBW volumes.^{20,21} On average, the TBW volume comprises approximately 63.3% of total body weight (0.5-0.6L per kg), of which 24.9% is located extracellularly and 38.4% intracellularly.²² More

Accepted for publication February 15, 2021.

From the *Department of Clinical Speech and Language studies, Trinity College Dublin, Dublin, Ireland; and the †The Library of Trinity College, Dublin, Ireland.

Address correspondence & reprint requests: Vasilis L. Georgalas, Trinity College Dublin, Phoenix House, 7-9 Leinster Street, Dublin 2, Dublin, Ireland. E-mail: georgalv@tcd.ie

Journal of Voice, Vol. ■■■, No. ■■■, pp. ■■■–■■■
0892-1997

© 2021 The Authors. Published by Elsevier Inc. on behalf of The Voice Foundation. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>)
<https://doi.org/10.1016/j.jvoice.2021.02.025>

TABLE 1.
Caffeine Consumption Safe Levels as Proposed by Health Canada

Age group/Condition	Amount of caffeine
Adults	≤400mg
Children (6-12years)	45-85mg
Adolescents (≥12years)	2.5mg per kg
Pregnant Women	<300mg

specifically, the extracellular volume contains 5% of plasma water and about 20% interstitial fluid.²²

Voice-related studies of hydration have been concerned with systemic and superficial hydration. Systemic hydration refers to adequate fluid located within body tissues and is predominantly achieved through water intake²³. Conversely, superficial (or surface) hydration is defined as the hydration of the surface of laryngeal mucosa that keeps the epithelial cells moisturized and lubricated, and is achieved via steam inhalation or increased environmental humidity.²⁴ Inadequate hydration can adversely affect vocal fold viscoelasticity, oscillation threshold and voice quality, since vocal folds are covered with a thin mucosal surface layer that has biomechanical and protective properties.²⁴

In vivo studies on excised animal larynges examined the physiology and biomechanics of the vocal folds after induced dehydration and rehydration challenges. The desiccation challenges utilized to dehydrate the excised animal larynges resulted in increased tissue stiffness with consequent increased phonation threshold pressure (PTP), vocal fold tension and viscosity alterations.²⁵⁻³⁰

Human studies have demonstrated that both systemic and superficial dehydration adversely affects phonation. Reduced hydration over a fasting period greatly affected maximum phonation time (MPT) and perceived phonatory effort (PPE) in males and females.^{30,31} Inadequate hydration also negatively impacted acoustic, perceptual and aerodynamic measures in professional singers following a 2-hour rehearsal.³² Three studies demonstrated that superficial dehydration due to low relative humidity (RH) adversely influenced PPE, PTP and other acoustic measures compared to moderate or high RH.³³⁻³⁵ Two more studies investigated the effects of oral breathing combined with various levels of RH and the results indicated that PPE and PTP measures were negatively affected, however the influence of low RH exacerbate the outcomes.^{36,37}

While increased water intake is a common recommendation by clinicians, especially if caffeinated beverages are consumed, the available evidence on the effects of caffeine on the body's hydration levels has yielded inconsistent results. Clinical advice is based on the anecdotal notion that caffeine increases diuresis, hence fluid balance within the body is affected. This systematic review seeks to summarize the available evidence about the effects of caffeine on voice-related measures in healthy participants.

MATERIALS AND METHODS

Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines were used in this systematic review.³⁸ A protocol was developed to reduce author bias and increase methodological quality.³⁹ The protocol was registered in the PROSPERO database⁴⁰ (PROSPERO ID: CRD42020196488).

Search strategy

Prior to search strategy development, scoping (preliminary) searches were utilized, using simple terms, in order to exclude the possibility of an existing SR in the same topic and to locate essential studies.⁴¹ Next, a search strategy was developed in collaboration with a subject librarian (IH). This search was piloted and refined to minimize irrelevant results. The search was devised to search for synonyms and related terms for: "healthy adults AND caffeine AND decaffeinated AND voice outcome measures". To ensure the search strategy's accuracy, the Peer Review of Electronic Search Strategy (PRESS) guidelines were used by the second reviewer (NK) to appraise the quality of the search and offer recommendations for change.⁴² Overall, the search strategy was deemed appropriate and no revisions were proposed. Six databases (PubMed, Cinahl Complete, Web of Science Core Collection, EMBASE, Cochrane Central, ProQuest Dissertation and Theses A&I) were searched for eligible studies in June 2020. The finalized search strategy consisted of 9 thesaurus (or subject heading) terms and 24 title & abstract terms. The 'human' filter option was used across databases to exclude animal studies. Otherwise, no filters or restrictions were used. A sample search strategy for the PubMed database is in Appendix. Backward citation chaining was also used. This refers to a system of identifying important articles through scanning the reference list of included studies.⁴³

Eligibility criteria

The PICOS framework⁴⁴ was used to guide study eligibility. Participants had to be vocally healthy without diagnosed voice disorders. No age or sex restrictions applied. Studies were eligible for inclusion only if the substance type (e.g. coffee, caffeine tablets, energy drinks) and dosage of caffeine consumed were reported. Although comparators were not a prerequisite for inclusion, if the study had a control group, the comparator could be a decaffeinated beverage, placebo or water. Outcomes of interest were any acoustic, aerodynamic, auditory-perceptual or self-perceptual voice measures.

Randomized control trials (RCTs) and non-randomized studies (NRS) were eligible irrespective of publication, date or language status.

Study selection

All retrieved studies were imported into Covidence and were automatically de-duplicated. The software de-duplicates the articles on the basis of author, title, year and volume of publication. Title and abstract, as well as full text screening were conducted independently by two reviewers (VG & NK) in

order to minimize the possibility of human error and bias.³⁹ Disagreements were resolved by consensus. In cases where potential disagreements could not be resolved, a third reviewer (CK) acted as an arbitrator by providing a casting vote.

Data analysis and synthesis

A data extraction form was designed to capture information about participants, study methods, outcome measures and results. Each section was revised and tailored to the review's needs by the principal investigator (VG). Prior to data extraction, the form was piloted on three randomly chosen included studies. Since the data and methodology of each study may vary, piloting was necessary in order to ensure that the data collection form could meet each study's requirements.³⁹ After piloting, the final data collection form included the following domains: population and setting, study methodology, participant information, study characteristics and experimental procedures, outcome measures.

Data were extracted independently by two reviewers (VG, NK). Missing data was addressed by the principal investigator (VG), who contacted authors in order to request information essential for this review. The extracted data were tabulated and summarized narratively, since a meta-analysis was not possible due to heterogeneity across studies.

Quality assessment

Risk of bias

Risk of bias was independently appraised by two reviewers (VG, NK). The Risk of Bias (RoB) tool by Cochrane Collaboration was selected in for RCTs. For NRS methodologies,

the Downs & Black (D&B) checklist was selected an appropriate tool that can be implemented in an array of study methodologies.⁴⁵ A modified version of D&B checklist was utilized.⁴⁶ This adapted version of D&B that was implemented in this SR consists of a 28-point scale compared to the original 32-point scale. The variation of the total score is a result of the altered scoring system of the last domain *Power*. In the original version by D&B the score in the domain *Power* ranges from 0 to 5 depending on the number of subjects allocated to each group. Conversely, in this adapted version the last question in the domain *Power* was scored on the basis of whether the study reported a power analysis or not. Thus, the score in this domain ranges from 0 to 1 (refer to Table 2 for a detailed description of the tool).

Levels of evidence

The Oxford Centre of Evidence Based Medicine (OCEBM) guide appraises the evidence level of a study based on type of evidence and study methodology.⁴⁷ This guide consists of seven questions. Each is a typical question a clinician might encounter when providing advice. For this review the question *What are the COMMON harms?* was chosen on the basis that caffeine is considered to be potentially harmful to the voice (Table 3). Other questions from the guide were not included as they were irrelevant. It was agreed by the review team to include pilot studies at Level 5 of evidence, as pilot studies, much like "mechanism-based reasoning" studies, explore the feasibility of larger scale RCTs and summarize information for future hypothesis testing.⁴⁸

TABLE 2.
Downs and Black Risk of Bias Tool With Adaptations by Hooper et al. (2008)

Domains	Scoring System	Score Range	Quality appraisal based on the overall score
Reporting	Yes=1, No=0,	0–28	Excellent → 26–28
External Validity	Unable to determine=0,		Good → 20–25
Internal Validity (Bias)	[Yes=2, Partially=1,		Fair → 15–19
Internal Validity (Confounding)	No=0 – used only in one item]		Poor → ≤ 14
Power			

TABLE 3.
OCEBM Treatment Harms Domain

What are the COMMON harms? (Treatment Harms)	
Level 1	Systematic review of randomized trials, systematic review of nested case-control studies, n-of-1 trial with the patient you are raising the question about, or observational study with dramatic effect
Level 2	Individual randomized trial or (exceptionally) observational study with dramatic effect
Level 3	Non-randomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)
Level 4	Case-series, case-control, or historically controlled studies
Level 5	Mechanism-based reasoning & pilot studies ^a

^a Pilot study designs were added by the review team.

RESULTS

Database search results

The total search results retrieved from all databases was $n=1818$. Following de-duplication, $n = 1443$ title and abstracts were screened and $n = 1435$ of them were rejected based on inclusion/exclusion criteria. A total of $n=8$ studies were deemed eligible for full-text screening. The full text of one study was not available online,⁴⁹ thus the primary investigator (VG) corresponded with the primary author. This study was excluded due to lack of response, therefore $n = 7$ studies were included for full-text review. During full-text review, two further studies were removed as they did not meet eligibility criteria. No disagreements between the reviewers occurred. A total of $n = 5$ studies were included for data extraction. The citation chaining method did not identify additional articles. The PRISMA flow diagram (Figure 1) illustrates the review phases.

Study characteristics

All eligible studies had non-randomized study designs.⁵⁰⁻⁵⁴ Three of five were pilot studies,^{50,51,53} while the remaining two were true experimental studies.^{52,54} The findings of the pilot study by Ahmed et al.⁵⁰ were published as a “Letter to the Editor”.

Sample size and participant characteristics

A total of $n=155$ healthy participants were recruited amongst the included studies. The age range was reported in four out of five studies and the overall range was 18 to 55 years,⁵¹⁻⁵⁴ with the most common age range between 18-35 years. Mean participant age was reported in only one study⁵³ and was 23. In another paper raw data were available, so mean age and standard deviation were computed as 22.7 ± 3.86 .⁵² Participant sex was reported in only four out of five studies and was 91% female, 9% male⁵¹⁻⁵⁴ (Table 4).

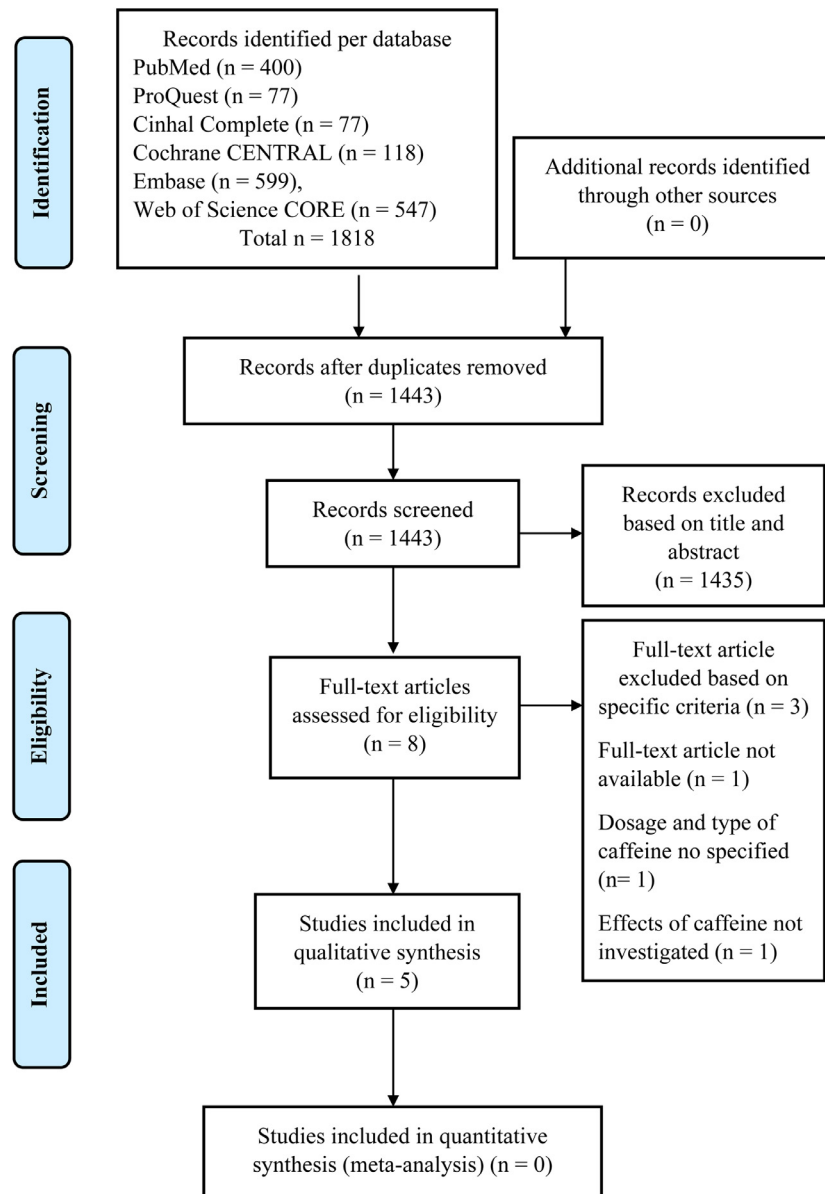


FIGURE 1. PRISMA flow-chart. Adapted from Moher et al. (2009)

TABLE 4.
Summary of Study and Participant Characteristics per Study

Reference	Study design	Participants			Sampling method	Eligibility criteria	
		N of participants	Gender ratio %	Age		Inclusion	Exclusion
Ahmed et al. 2012	Pilot, repeated measures, sham-controlled, experimental study	N=25	N/R	N/R ^a	N/R	No diagnosis of voice disorder, no oral cavity disorders	Recent upper respiratory tract infection, hypothyroidism, renal or blood pressure problems
Akhtar et al. 1999	Pilot, repeated measures, one group study	N=8	Male: 50% Female: 50%	Mean: N/R Median: N/R Range: 27-55	N/R	No recent diagnosis of voice disorders	Diagnosis of systemic illnesses especially respiratory or cardiovascular
Erickson-Levendoski & Sivansankar, 2011	Prospective, double-blinded, sham-controlled repeated measures experimental study	N=16	Male: 50% Female: 50%	Mean: 22.7 Median: 20 Range: 18-32	N/R	Perceptually normal speech and voice, no hearing problems	Respiratory disease, reflux, smoking, prescription medication except oral contraceptives
Franca & Simpson, 2013	Pilot, repeated measures, experimental study	N=48	Male: 0% Female: 100%	Mean: 23 Median: - Range: 18-35	N/R	Self-reported normal voice	High blood pressure, coronary disease, self-reported discomfort while consuming caffeine
Franca et al., 2013	Prospective, repeated measures, experimental design	N=58	Male: 0% Female: 100%	Mean: N/R Median: - Range: 18-35	N/R	Self-reported normal voice, general good health	High blood pressure, coronary disease

^a N/R= Not reported

With regards to participant eligibility, perceptual or self-reported normal voice with absence of voice pathology was the main inclusion criterion in every study. In the majority of the studies, subjects were excluded if high blood pressure and coronary disease were reported.^{50,51,53,54} Reflux symptoms, smoking and medication (except oral contraceptives) were exclusion criteria in one study.⁵² Lastly, three studies deemed participants ineligible for inclusion on the basis of self-reported or diagnosed respiratory disorders.⁵⁰⁻⁵² (Table 4)

Experimental procedures

Variation was noted in the experimental procedures employed in each study (Table 5). With regards to the intervention, the type and dose of the caffeine varied amongst the studies. The participants in three of the studies consumed caffeine tablets^{51,53,54} while in the rest of the studies subjects ingested coffee.^{50,52} The milligrams (mg) of caffeine ingested by participants in included studies ranged from 100mg to 480mg. The comparator intervention consisted of placebo, water, decaffeinated coffee and no intervention at all. While all studies had pre-caffeine baseline measures, only four of them collected voice measures on the same day. The duration of the study procedures varied from two hours to two days, though duration was not reported in two studies^{53,54}. Abstention from caffeine to better control experimental procedures was used in four out of five studies, but with varying instructions.

Outcome measures

The primary outcome measures utilized in the included studies were aerodynamic, acoustic and perceptual. Acoustic measures like jitter and shimmer were the most frequently reported outcomes, while perceptual measures were utilized in only one study. Perceptual and acoustic measures were obtained under different conditions, such as sustained “ah” sound, singing, reading or speaking. PTP and aerodynamic airflow were the only aerodynamic measures collected. PTP and airflow outcomes were obtained while the subjects repeatedly uttered the syllables /pi/ or /pa/ respectively. Secondary outcome measures were considered in one study, where the participants were requested to rate their vocal effort using a visual analog scale⁵². Overall, all studies reported non-significant effects of caffeine on voice-related measures ($P \geq 0.05$). To be noted, only two studies reported the exact p values for each outcome measure^{53,54}. One study reported subtle changes in irregularity of fundamental frequency, however the authors attributed these irregularities to individual characteristics⁵¹. Detailed description of the results is provided in Table 6.

Methodological quality assessment [Table 7 near this section]

Risk of bias appraisal was conducted using the D&B checklist. Overall, none of the studies was deemed of ‘Good’ or

‘Excellent’ quality (Table 7). Three of the studies were rated as ‘Fair’ and the remaining two as ‘Poor’ quality. *Power*, *External Validity* and *Internal Validity-Confounding* domains presented the highest risk of bias between domains in all studies. Conversely, low risk of bias was noted in the *Reporting* domain in three studies, however in the rest of the studies the risk of bias in the same domain was unclear.

External Validity was at high risk due to lack of information regarding the sampling method, nonrepresentative population and non-representative experimental conditions. Random allocation of participants into intervention or control groups was not performed in any of the experiments and potential confounders were not provided, thus high risk of bias was observed in the *Internal Validity-Confounding* domain. Lastly, none of the studies presented evidence of power analysis. Despite the fact that a power analysis was not conducted, it can be deduced that the power level was reduced as it was influenced by small sample sizes, which may have increased the chance for a Type II error.

Using the OCEBM Levels of Evidence, the two experimental studies were placed at Level 3 of evidence, since they were non-randomized experimental studies. Pilot studies were automatically assigned in the Level 5 of evidence, as their study design explores the feasibility of a larger scale studies and do not provide substantial evidential information.

DISCUSSION

The aim of this study was to identify and critically appraise the available evidence regarding the potential effects of caffeine on voice-related measures. Due to small number of included studies, lack of methodological integrity and high risk of bias, the evidence regarding the effects of caffeine on phonation is unreliable. This review cannot therefore provide robust advice about the effects of caffeine on voice.

Methodological considerations

Caffeine is known to have a potentially systemic dehydrating effect. The degree of localised dehydration within the vocal folds is however unknown. Vocal fold dehydration can lead to aberrant voice quality through mechanisms such as reduction in vocal fold lubrication, reduced oscillation and increased risk of trauma through vocal fold collision.²³ These findings have been confirmed in a canine model, where dehydrated vocal folds were found to be stiffer and more viscous.²⁵ The studies in this systematic review are thus based on the hypothesis that systemic dehydration may induce vocal fold dehydration, which would manifest as abnormal voice production due to the factors above.

All included studies had a repeated measures experimental design in order to demonstrate the effects of caffeine intake. This is considered acceptable to measure the cause-and-effect relationships between independent and dependent variables between groups.⁵⁵ Three of these however employed the pre-post design in a pilot study methodology. Pilot studies provide non-evidential information, as they are

TABLE 5.
Overview of experimental Procedures in Included Studies

Reference	Intervention	Comparator	Duration	Absorption of caffeine	Hydration status measurements	Environmental parameters
Ahmed et al. 2012	Type: Coffee Dosage: 400mg	Type: Decaffeinated coffee Dosage: Not clearly defined	2 days	N/R	N/R	N/R
Akhtar et al. 1999	Type: Caffeine tablets Dosage: 250mg (5 × 50mg)	None	Approximately 2 hours	N/R Blood tests employed to ensure absorption	N/R	N/R
Erickson-Levendoski & Sivansankar, 2011	Type: Coffee Dosage: 480mg (2 × 240mg)	Type: Decaffeinated coffee Dose: 24mg 2 coffees consumed (12 × 2=24mg)	Approximately 5 hours	2.5-3hours	N/R	Environmental humidity control during measurements (70% ± 6%)
Franca & Simpson, 2013	Type: Caffeine tablet Dosage: 200mg	None	N/R	30mins	N/R	N/R
Franca et al., 2013	Type: Caffeine tablet Dosage: 100mg	Type: Placebo (tablet) Dose: 0mg	N/R	30mins	N/R	N/R

TABLE 6.
Overview of Outcome Measures and Statistical Findings

Reference	Outcome measures		Measurement Conditions	Statistical Findings
	Primary outcomes	Secondary outcomes		
Ahmed et al. 2012	<i>Acoustic:</i> Jitter, Shimmer, f0 <i>Auditory-Perceptual:</i> GRBAS	-	<i>Acoustic:</i> Sustained “ah” sound <i>GRBAS:</i> Grandfather Passage	N/R
Akhtar et al. 1999	F0 irregularity (using a laryngograph)	-	Free speech, reading, singing (“Happy Birthday”)	Effects of caffeine on f0 non-significant ($p \geq 0.05$) Effects of caffeine on PTP (10 th –80 th) and PPE non-significant ($p \geq 0.05$)
Erickson-Levendoski & Sivansankar, 2011	<i>Aerodynamic:</i> Phonation threshold pressure (PTP)	Perceived phonatory effort (PPE)	<i>PTP:</i> Measures obtained in two different pitches: Sustained “ah” sound obtained at 10% and 80% of the participant’s pitch range <i>PPE:</i> Self-rated vocal effort after singing (“Happy Birthday”) using a visual analog scale	Differences between caffeinated and decaffeinated conditions non-significant ($p \geq 0.05$)
Franca & Simpson, 2013	<i>Aerodynamic:</i> Aerodynamic measure of airflow <i>Acoustic:</i> Jitter, Shimmer, Sound Pressure Level	-	<i>All measures:</i> Sustained “ah” sound, x3 times repetition “Pa” syllable produced x3 times	Effects of caffeine consumption with or without water on voice measures Jitter: $p = 0.985$ Shimmer: $p = 0.833$ SPL: $p = 0.513$ Airflow: $p = 0.742$ } $P > 0.05$
Franca et al., 2013	<i>Aerodynamic:</i> Aerodynamic measure of airflow <i>Acoustic:</i> Jitter, Shimmer, Sound Pressure Level	-	<i>All measures:</i> Sustained “ah” sound, x3 repetitions “Pa” syllable produced x3 times	Effects of caffeine on voice measures Jitter: $p = 0.88$ Shimmer: $p = 0.35$ SPL: $p = 0.46$ Airflow: $p = 0.42$ } $P > 0.05$

TABLE 7.
Risk of Bias Overall Score per Domain for Each Study

Reference	Reporting	External Validity	Internal Validity Bias	Internal Validity Confounding	Power	Overall Score & Quality
Ahmed et al. 2012	7	0	3	1	0	11/28 'Poor'
Akhtar et al. 1999	7	0	2	1	0	10/28 'Poor'
Erickson-Levendoski & Sivansankar, 2011	9	0	7	2	0	18/28 'Fair'
Franca & Simpson, 2013	9	0	4	2	0	15/28 'Fair'
Franca et al., 2013	9	0	4	2	0	15/28 'Fair'

designed to explore the feasibility of larger-scale studies and not answer hypotheses or draw firm conclusions.⁴⁸

The recruited subjects were not representative of the entire population, hence a higher risk of selection bias is noted. Indeed, the studies' participants were predominantly females, and the age range in three out of five studies was between 18 and 35 years old. Therefore, the demographic characteristics of the subjects do not represent the true values of the population and differences based on sex could not be drawn.

Sample size and power level are bidirectionally correlated; the bigger the sample the greater the power level. Since the majority of the studies recruited a relatively small number of participants, the power of the studies was reduced, as it is influenced by the sample size. Consequently, the probability of detecting any effects is reduced and the probability of a Type II error is increased.⁵⁶ Another limitation that adversely influenced the generalizability and applicability of the evidence is the type of intervention implemented in the studies. The participants of the intervention groups predominantly consumed caffeinated tablets, a source of caffeine that does not reflect a realistic caffeine source, as the average populations' source of caffeine is coffee, tea and soft drinks.⁵⁷ This disimproved the ecological validity of the studies.

Experimental considerations

The findings suggest that voice production was not adversely affected by caffeine consumption. This non-significant outcome could be attributed to a variety of factors. The majority of the studies utilized moderate caffeine dosages, which are considered safe for the human body.¹⁷ Another factor that should be taken into account is the caffeine absorption rate. Caffeine absorption is completed in approximately 50 minutes; however, caffeine can be detected in blood plasma within approximately 35 minutes.⁵⁸ In two of the studies, voice measures were collected 30 minutes following caffeine consumption.^{53,54} Although theoretically 30 minutes are sufficient for the

caffeine to be detected in blood plasma, absorption and metabolism of caffeine varies amongst subjects depending on individual characteristics like sex and pharmacokinetics.¹⁵ Since the presence of caffeine in blood was not objectively measured, it is impossible to determine whether the caffeine was completely absorbed within the 30 minutes allotted. One study did however use blood tests for detection of caffeine plasma and the values of caffeine concentrations in blood increased following caffeine consumption.⁵¹ Following oral intake, caffeine's distribution is completed within 20 minutes through the biological membranes.⁵⁹ However, regular consumption of higher caffeine dosages can increase the distribution and excretion rate of caffeine.^{59,60} Participants' caffeine consumption habits were not reported in any of studies, thus differences in distribution and excretion rates may have been occurred. Caffeine's metabolic rate is significantly increased for smokers, a factor that was taken into account in only one study^{15,52}.

Caffeine tolerance is another variable that could have influenced outcomes. Information about the average daily caffeine intake of participants was not provided in any of the studies. Various studies have examined the tolerance levels and the effects of caffeine on habitual and non-habitual coffee drinkers. It is reported that repeated caffeine consumption can induce tolerance, a factor that mitigates caffeine's effects, such as diuresis, increased alertness, sleep deprivation and reduced sense of fatigue^{2,7}. Tolerance can be induced in approximately 10 days, however the degree of tolerance varies amongst individuals, as it depends on individual characteristics (eg, sex, age).¹⁴

Objective assessment of hydration status was not employed in any of the studies (eg, bioelectrical impedance analysis, urine samples). Assessment of hydration could provide an insight on possible systemic hydration status following caffeine consumption, since hydration levels vary amongst individuals. It should be noted though that the investigators in two of the studies instructed participants to avoid liquids 12 hours prior to the experimental procedures,^{53,54} in an attempt to create equal baseline levels of hydration. Since an objective hydration assessment was not

employed, it was impossible to determine whether the participants adhered to the investigators' instructions or whether equal baseline hydration levels were achieved. The effects of environmental humidity and its effects on superficial hydration of the vocal folds were acknowledged in one study where the investigators adjusted the RH to moderate levels (70% RH).⁵²

Overall, a few potential confounders were taken into account (eg, menstrual cycle, smoking), however none of the experiments described comprehensive experimental control by listing and exhaustively controlling potential confounders.

Implications for clinical practice

Physicians and speech and language therapists advise patients to refrain from caffeine consumption on the assumption that caffeine intake induces diuresis, which lowers fluid balance. Such imbalances can induce dehydration which can be detrimental to phonation. The implied transitive relationship must however be supported empirically. This systematic review focused on the impact of caffeine on voice in healthy adults and found inconclusive evidence about any deleterious effects. Due to lack of methodological quality and risk of bias, the outcomes of this review do not provide robust evidence regarding the potential adverse effects of caffeine on phonation. Thus, clinicians should be cautious when counselling patients to refrain from caffeine, as clinical recommendations cannot be supported. Health professionals could refer to published guidelines¹⁷ regarding the safe values of caffeine consumption to advise patients to moderate caffeine intake if uncertain.

This is not to say that counselling patients to reduce caffeine intake is totally without merit. Clinicians should consider the other relationships between caffeine with voice. For instance, caffeine intake might exacerbate existing disorders such as laryngopharyngeal reflux, making avoidance appropriate in some cases.⁶¹ *In vivo* studies⁶² have also demonstrated that caffeine can interfere with circadian rhythms, altering sleep and/or wake cycles. While a systemic effect and not one isolated to the vocal folds, this may cause those with voice difficulties not to obtain refreshing sleep, which could exacerbate feelings of stress. Despite the fact that caffeine has been associated with lower risk of depression, caffeine consumption might adversely affect people with mental health issues. Caffeine has been positively associated with an increased risk for deterioration of anxiety symptoms and higher risk of relapse episodes.⁶³ SLTs working with patients with mental health issues should consider the effects of caffeine consumption on these patients, since no specific guidelines have been published for safe dosages of caffeine consumption amongst people with mental and psychiatric health issues.⁶³

Directions for future research

The low quantity and quality of the studies that were identified highlights the need for further, more robust research.

Five studies were deemed eligible for inclusion, but all were characterized by methodological flaws and high risk of bias rendering the results unreliable. The methodological limitations of the included studies that are described in this review could however provide the basis for more robust investigations of the effects of caffeine on phonation in the future.

RCTs utilizing randomization (eg, allocation, sampling) and blinding could ensure that cause-and-effect relationships may be attributed to the implemented intervention. However, RCTs are not always feasible, so carefully designed true experimental studies can also yield evidence that can be internally valid. Future investigations must focus on controlling extraneous factors that potentially influence the outcomes. Controlling for all potential confounders might not be possible, thus it is essential to include an objective hydration assessment method in order to determine whether caffeine has an influence on the hydration status of the participants and to utilize blood or urine tests to measure caffeine absorption.

A recent review by Wikoff et al.⁶ reviewed data for potential adverse effects of caffeine consumption and the results showed that low to moderate doses (up to 400mg) of caffeine do not negatively affect the body. Thus, greater dosages of caffeine (>400mg) should be used to increase the possibility of detecting effects on voice quality and to reflect the fact that some individuals consume caffeine above recommended amounts. Future research should examine caffeine concentrations in the medium in which they are served. For instance, weaker coffee (higher water to caffeine ratio) might promote hydration and mitigate the dehydrating effects of caffeine as compared to another study that uses stronger coffee (lower water to caffeine ratio). In addition, control over environmental confounders such as humidity is advisable, as superficial hydration of the vocal fold tissues is achieved through higher RH, a factor that might influence outcomes.

Prospective studies should recruit an adequate number of participants that will be representative of the entire population. A large sample size would ensure adequate power levels, so the probability of detecting potential effects would be increased. Additionally, effects should be investigated amongst different age groups and sexes in order to render outcomes more generalizable and representative. Another factor that could be taken into consideration is the long-term effects of caffeine consumption on phonation. Future prospective longitudinal studies could shed light to the long-term effects of caffeine intake on phonation, rendering them more representative of real-life situations, where caffeine is consumed regularly and not in isolation.

It should be noted that no studies that investigated the effects of caffeine in children under-18 years of age were identified, highlighting the lack of research and evidence in this area. In terms of occupation, future research should also investigate the potentially adverse effects of caffeine on occupational and professional voice users, a population that is particularly prone to voice disorders.

This systematic review focused on the impact of caffeine on voice in healthy adults and found inconclusive evidence about any deleterious effects. Clinicians working with those who have voice disorders may be interested in how caffeine affects those with dysphonia, since dysphonic individuals could be at increased risk of harm due to their underlying pathology. Even modest dehydration of the vocal folds in such individuals could exacerbate voice difficulties and it would be illuminating to identify whether caffeine might contribute to this. Any researchers engaging in such studies would however need to carefully consider the ethical difficulties of potentially exposing participants to further vocal harm, in addition to controlling for anatomical, physiological and biomechanical variation between different voice disorders.

Limitations

Due to small number of studies, poor quality of data and lack of homogeneity amongst the studies a meta-analysis was not conducted, hence a quantitative analysis and presentation of the findings was not possible.

Although sample sizes were small in a few studies, and although this most likely had a detrimental effect on statistical power, two of the studies utilized relatively high sample sizes. Since a power analysis was not computed, a precise estimation of the ability of the included studies to establish an effect could not be determined.

CONCLUSION

The findings of the present review cannot provide robust evidence regarding the effects of caffeine on voice-related measures. Since no firm conclusions can be elicited to guide clinical practice, clinicians should be cautious when recommending caffeine abstinence to patients.

The results of this review demonstrate the lack of research in the field and the necessity to inform evidence-based practice through reliable and valid outcomes. Future research should recruit a more representative sample and employ robust experimental procedures.

DECLARATION OF INTEREST STATEMENT

No potential conflict of interest was reported by the authors.

APPENDICES

Database Electronic Search Example: PubMed

[MeSH Terms] & [Title/Abstract]

1. Dehydration [MeSH] OR Beverage [MeSH] OR Caffeine [MeSH] OR Coffee [MeSH]
2. coffee [Title/Abstract] OR caffeine* [Title/Abstract] OR decaffeain* [Title/Abstract] OR de-caffein* [Title/Abstract] OR de caffein* [Title/Abstract] OR beverage* [Title/Abstract] OR hydrat* [Title/Abstract] OR rehydrat*[Title/Abstract] OR re-hydrat* [Title/Abstract] OR re hydrat* [Title/Abstract] OR

dehydrat* [Title/Abstract] OR de-hydrat* [Title/Abstract] OR de hydrat*[Title/Abstract]

3. 1 OR 2
4. Phonation [MeSH] OR Larynx [MeSH] OR Voice [MeSH] OR Voice Disorders [MeSH] OR Speech Acoustics [MeSH]
5. “speech acoustic*” [Title/Abstract] OR “perturbation measure*” [Title/Abstract] OR “aerodynamic measure*” [Title/Abstract] OR “acoustic measure*” [Title/Abstract] OR “perceived phonatory effort” [Title/Abstract] OR voice [Title/Abstract] OR phonation [Title/Abstract] OR larynx [Title/Abstract] OR larynges [Title/Abstract] OR laryngeal [Title/Abstract] OR vocal [Title/Abstract]
6. 4 OR 5
7. 3 AND 6

(phonation[MeSH Terms] OR larynx[MeSH Terms] OR voice disorders[MeSH Terms] OR voice[MeSH Terms] OR speech acoustics[MeSH Terms] OR speech acoustic*[Title/Abstract] OR perturbation measure*[Title/Abstract] OR aerodynamic measure*[Title/Abstract] OR acoustic measure*[Title/Abstract] OR perceived phonatory effort[Title/Abstract] OR voice[Title/Abstract] OR phonation[Title/Abstract] OR larynx[Title/Abstract] OR larynges[Title/Abstract] OR laryngeal[Title/Abstract] OR vocal[Title/Abstract]) AND (dehydration[MeSH Terms] OR beverage [MeSH Terms] OR caffeine[MeSH Terms] OR coffee [MeSH Terms] OR coffee[Title/Abstract] OR caffeine*[Title/Abstract] OR decaffeain*[Title/Abstract] OR de-caffein*[Title/Abstract] OR de caffein* [Title/Abstract] OR beverage*[Title/Abstract] OR hydrat*[Title/Abstract] OR rehydrat*[Title/Abstract] OR re hydrat* [Title/Abstract] OR dehydrat*[Title/Abstract] OR de-hydrat*[Title/Abstract] OR de hydrat*[Title/Abstract])

REFERENCES

1. Carrillo JA, Benitez J. Clinically significant pharmacokinetic interactions between dietary caffeine and medications. *Clin Pharmacokinetics*. 2000;39:127–153.
2. Armstrong LE. Caffeine, body-fluid electrolyte balance, and exercise performance. *Int J Sport Nutrition and Exercise Metabolism*. 2002;12:189–206.
3. Heckman MA, Weil J, de Mejia EG. Caffeine (1, 3, 7-trimethylxanthine) in foods: A comprehensive review on consumption, functionality, safety, and regulatory matters. *J Food Science*. 2010;75:77–87.
4. Poole R, Kennedy OJ, Roderick P, et al. Coffee consumption and health: Umbrella review of meta-analyses of multiple health outcomes. *BMJ (Clin res ed.)*. 2017;359:j5024.
5. Gonzalez de Mejia E, Ramirez-Mares MV. Impact of caffeine and coffee on our health. *Trends in Endocrinol & Metabolism: TEM*. 2014;25:489–492.
6. Wikoff D, Welsh BT, Henderson R, et al. Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children. *Food and Chemical Toxicol*. 2017;109(Pt 1):585–648.
7. Maughan RJ, Griffin J. Caffeine ingestion and fluid balance: A review. *J Human Nutrition and Dietetics*. 2003;16:411–420.

8. Zhang Y, Coca A, Casa DJ, et al. Caffeine and diuresis during rest and exercise: a meta-analysis. *J Science and Medicine in Sport*. 2015;18:569–574.
9. Bird ET, Parker BD, Kim HS, et al. Caffeine ingestion and lower urinary tract symptoms in healthy volunteers. *NeuroUrol and Urodynamics*. 2005;24:611–615.
10. Neuhäuser-Berthold M, Beine S, Verwied SC, et al. Coffee consumption and total body water homeostasis as measured by fluid balance and bioelectrical impedance analysis. *Ann of Nutr and Metabolism*. 1997;41:29–36.
11. Wemple R, Lamb D, McKeever K. Caffeine vs caffeine-free sports drinks: Effects on urine production at rest and during prolonged exercise. *Int J Sports Med*. 1997;18:40–46.
12. Fiala KA, Casa DJ, Roti MW. Rehydration with a caffeinated beverage during the nonexercise periods of 3 consecutive days of 2-a-day practices. *Int J of Sport Nutr and Exercise Metabolism*. 2004;14:419–429.
13. Silva AM, Júdice PB, Matias CN, et al. Total body water and its compartments are not affected by ingesting a moderate dose of caffeine in healthy young adult males. *Applied Physiology, Nutrition, and Metabolism*. 2013;38:626–632.
14. Grandjean AC, Reimers KJ, Bannick KE, et al. Effect of caffeinated, non-caffeinated, caloric and non-caloric beverages on hydration. *Journal of the American College of Nutrition*. 2000;19:591–600.
15. Nehlig A. Interindividual differences in caffeine metabolism and factors driving caffeine consumption. *Pharmacological Rev*. 2018;70:384–411.
16. Temple JL, Bernard C, Lipshultz SE, et al. The safety of ingested caffeine: a comprehensive review. *Frontiers in Psychiatry*. 2017;8:80.
17. Government of Canada, Health Canada. Caffeine in food. [2012, February 16]. Available from: <https://www.canada.ca/en/health-canada/services/food-nutrition/food-safety/food-additives/caffeine-foods/foods.html>. [Accessed 15 April 2020].
18. Nawrot P, Jordan S, Eastwood J, et al. Effects of caffeine on human health. *Food Additives and Contaminants*. 2003;20:1–30.
19. Hartley NA, Thibeault SL. Systemic hydration: Relating science to clinical practice in vocal health. *J Voice*. 2014;28:652.e1–652.e20.
20. Kavouras SA. Assessing hydration status. *Current Opinion in Clin Nutrition and Metabolic Care*. 2002;5:519–524.
21. Manz F, Wentz A. 24-h hydration status: Parameters, epidemiology and recommendations. *Eur J Clinical Nutrition*. 2003;57(Suppl. 2):S10–S18.
22. Armstrong LE. Hydration assessment techniques. *Nutrition Rev*. 2005;63:40–54.
23. Alves M, Krüger E, Pillay B, et al. The effect of hydration on voice quality in adults: a systematic review. *J Voice*. 2017;33:125.e13–125.e28.
24. Leydon C, Sivasankar M, Falciglia DL, et al. Vocal fold surface hydration: a review. *J Voice*. 2009;23:658–665.
25. Chan RW, Tayama N. Biomechanical effects of hydration in vocal fold tissues. *Otolaryngol-Head and Neck Surg*. 2002;126:528–537.
26. Haji T, Mori K, Omori K, et al. Experimental studies on the viscoelasticity of the vocal fold. *Acta Oto-Laryngologica*. 1992;112:151–159.
27. Hemler RJB, Wieneke GH, Lebacqz J, et al. Laryngeal mucosa elasticity and viscosity in high and low relative air humidity. *European Archives of Oto-Rhino-Laryngology*. 2001;258(3):125–129.
28. Jiang J, Verdolini K, Jennie N, et al. Effects of dehydration on phonation in excised canine larynges. *Annals of Otolaryngology, Rhinology & Laryngology*. 2000;109:568–575.
29. Witt RE, Regner MF, Tao C, et al. Effect of dehydration on phonation threshold flow in excised canine larynges. *Ann Otol, Rhinol & Laryngol*. 2009;118:154–159.
30. Hamdan AL, Sibai A, Rameh C. Effect of fasting on voice in women. *J Voice*. 2007;21:495–501. <https://doi.org/10.1016/j.jvoice.2006.01.009>.
31. Hamdan A, Ashkar J, Sibai A, et al. Effect of fasting on voice in males. *Am J Otolaryngol*. 2011;32:124–129.
32. van Wyk L, Cloete M, Hattingsh D, et al. The effect of hydration on the voice quality of future professional vocal performers. *J Voice*. 2017;31:111.e29–111.e36.
33. Hemler RJB, Wieneke GH, Dejonckere PH. The effect of relative humidity of inhaled air on acoustic parameters of voice in normal subjects. *J Voice*. 1997;11:295–300.
34. Levendoski EE, Sundarrajan A, Sivasankar MP. Reducing the negative vocal effects of superficial laryngeal dehydration with humidification. *Ann Otol, Rhinol & Laryngol*. 2014;123:475–481.
35. Verdolini K, Titze IR, Fennell A. Dependence of phonatory effort on hydration level. *J Speech Language and Hearing Res*. 1994;37:1001.
36. Sivasankar M, Erickson E, Schneider S, et al. Phonatory effects of airway dehydration: Preliminary evidence for impaired compensation to oral breathing in individuals with a history of vocal fatigue. *J Speech Language and Hearing Res*. 2008;51:1494–1506.
37. Sivasankar MP, Erickson-Levendoski E. Influence of obligatory mouth breathing, during realistic activities, on voice measures. *J Voice*. 2012;26:813.e9–813.e13.
38. Moher D, Liberati A, Tetzlaff J, et al. PRISMA Group Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med*. 2009;6: e1000097.
39. Higgins JP, Thomas J, Chandler J, et al. *Cochrane handbook for systematic reviews of interventions*. 2nd ed Chichester, UK: John Wiley & Sons; 2019.
40. PROSPERO: International prospective register of systematic reviews. 10/02/2020. Unique Identification number CRD42020196488, still under assessment. Available from: https://www.crd.york.ac.uk/prospéro/display_record.php?RecordID=196488.
41. Boland A, Cherry GM, Dickson R. *Doing a systematic review: A student's guide*. 2nd ed. London, UK: Sage Publications Ltd; 2017.
42. McGowan J, Sampson M, Salzwedel DM, et al. Press peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiol*. 2016;75:40–46.
43. Cribbin TF. Citation chain aggregation: An interaction model to support citation cycling. *Proceedings of the 20th ACM Int Conference on Information and Knowledge Management – CIKM*. 2011;11:2149–2152.
44. CRD (Centre for reviews and dissemination). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. York, UK: York Publishing Services Ltd; 2009.
45. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions. *J Epidemiol & Community Health*. 1998;52:377–384.
46. Hooper P, Jutai JW, Strong G, et al. Age-related macular degeneration and low-vision rehabilitation: a systematic review. *Canadian J Ophthalmol. J Canadien d'Ophthalmologie*. 2008;43:180–187.
47. Howick J, Chalmers I, Glasziou P, et al. Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document). *Oxford Centre for Evidence-Based Med*. 2011. Available from: <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/explanation-of-the-2011-ocbm-levels-of-evidence>. Accessed 1 April 2020.
48. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. *J Psychiatric Res*. 2011;45:626–629.
49. Franca, M. C. Interaction of caffeine and hydration on voice (Clinical Trial Protocol). 2011; *ClinicalTrials.gov ID: NCT01435837*.
50. Ahmed S, Coomber S, Chetwood T. A pilot randomised control trial: The effects of decaffeinated drinks on voice quality. *Clin Otolaryngol*. 2013;37:428–431.
51. Akhtar S, Wood G, Rubin JS, et al. Effect of caffeine on the vocal folds: a pilot study. *The J Laryngol & Otol*. 1999;113:341–345.
52. Erickson-Levendoski E, Sivasankar M. Investigating the effects of caffeine on phonation. *J Voice*. 2011;25:e215–e219.
53. Franca MC, Simpson KO. Effects of the interaction of caffeine and water on voice performance. *Communication Disorders Quarterly*. 2013;35:5–13.
54. Franca MC, Simpson KO, Schuette A. Effects of caffeine on vocal acoustic and aerodynamic measures of adult females. *CoDAS*. 2013;25:250–255.
55. Bowling A. *Research methods in health: Investigating health and health services*. Buckingham, UK: Open University Press; 2009.

56. Cadeddu MO, Farrokhyar F, Thoma A, et al. Users' guide to the surgical literature: how to assess power and sample size. *Canadian J Surg.* 2008;51:476–482.
57. Verster JC, Koenig J. Caffeine intake and its sources: a review of national representative studies. *Critical Rev in Food Science and Nutrition.* 2017;58:1250–1259.
58. Flaten MA, Aasli O, Blumenthal TD. Expectations and placebo responses to caffeine-associated stimuli. *Psychopharmacology.* 2003;169:198–204.
59. Nieber K. The impact of coffee on health. *Planta Medica.* 2017;83:1256–1263.
60. Magkos F, Kavouras SA. Caffeine use in sport, pharmacokinetics in man, and cellular mechanisms of action. *Critical Rev in Food Science and Nutrition.* 2005;45:535–562.
61. Martinucci I, de Bortoli N, Savarino E, et al. Optimal treatment of laryngopharyngeal reflux disease. *Therapeutic Advances in Chronic Disease.* 2013;4:287–301.
62. Burke TM, Markwald RR, McHill AW, et al. Effects of caffeine on the human circadian clock in vivo and in vitro. *Science Translational Medicine.* 2015;7. 305ra146–305ra146.
63. Temple JL, Bernard C, Lipshultz SE, et al. The safety of ingested caffeine: A comprehensive review. *Frontiers in Psychiatry.* 2017;8:19.