Systematic review

Blijleven, Esther E; Fuchten, Denise; Dullaart, Max J; Stokroos, Robert J; Thomeer, Henricus G X M; Wegner, Inge

Published in:
Laryngoscope investigative otolaryngology

DOI:
10.1002/lio2.1108

Publication date:
2023

Received: 26 April 2023 | Revised: 7 June 2023 | Accepted: 25 June 2023

DOI: 10.1002/lio2.1108

REVIEW

Laryngoscope Investigative Otolaryngology

Systematic review: Validity, reliability, and diagnostic accuracy of the electrogustometer

Esther E. Blijleven MD[1,2] | Denise Fuchten MSc[1,2] | Max J. Dullaart MD[1] | Robert J. Stokroos MD, PhD[1,2] | Henricus G. X. M. Thomeer MD, PhD[1,2] | Inge Wegner MD, PhD[3]

[1]Department of Otorhinolaryngology—Head and Neck Surgery, University Medical Center Utrecht, Utrecht, The Netherlands

[2]University Medical Center Utrecht Brain Center, Utrecht University, Utrecht, The Netherlands

[3]Department of Otorhinolaryngology—Head and Neck Surgery, University Medical Center Groningen, Groningen, The Netherlands

Correspondence
Esther E. Blijleven, Department of Otorhinolaryngology—Head and Neck Surgery, University Medical Center Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands.
Email: e.e.blijleven-2@umcutrecht.nl

Abstract

Objective: What are the electrogustometer's (EGM) validity, reliability, and diagnostic accuracy in assessing taste sensation in adults compared to other taste tests?

Data Sources: PubMed Medline, Elseviers's Embase, and the six databases of Cochrane Library.

Methods: We conducted a systematic search on December 20, 2022, consisting of synonyms for EGM. We considered randomized controlled trials and observational studies with original data for inclusion if they included adults who underwent electrogustometry. Articles were excluded if no analysis regarding validity, reliability, or diagnostic accuracy had been performed or if these analyses could not be performed with the published data.

Results: Nineteen articles discussing 18 studies were included for data extraction. The included studies carry a high risk of bias. Overall, the association between a variety of reference taste tests and EGM was moderate or weak with correlation coefficients ranging from −0.51 to 0.40 with one outlier of −0.74 found in one study correlating EGM and taste solutions. Test–retest reliability was good with reported correlation coefficients between 0.78 and 1.0. The sensitivity, specificity, PPV, and NPV of EGM in identifying abnormal taste function varied widely between the four studies on diagnostic accuracy.

Conclusion: The included studies in this review lack the required standards regarding study design to draw firm conclusions about the validity, reliability, and diagnostic accuracy of the EGM. Future research is needed to assess these measurement properties. Based on the reported results, we would not recommend using the EGM as a screening test for taste disturbance in clinical practice.

Level of Evidence: NA.

KEYWORDS

diagnostic accuracy, electrogustometry, reliability, systematic review, taste sensation, validity

1 | INTRODUCTION

Since 1958, the electrogustometer (EGM) has been used as a clinical tool for measuring electrical taste thresholds of the three gustatory nerves: the chorda tympani, glossopharyngeal and greater petrosal nerves.[1,2] Over the years, several EGM models have been developed. All models produce microamp-level anodic currents to the tongue, causing sour or metallic taste sensations. There are two hypotheses regarding the mechanism of electrogustometry. Based on the electrolyte chemical hypothesis, the anodic current creates an acidic solution through electrolysis, which innervates the taste receptor cells. Based on the direct effect hypothesis, the electrical taste arises from the anodic current action on nerve fibers or taste receptor cells.[3]

The methods for assessing patients with taste disorders have not been fully standardized.[4] Quantitative assessment of gustatory function with the taste strip test (TST) and filter paper disk (FPD) are established methods.[5] The advantages of using an EGM in comparison to the TST and FPD are that only a short test time is required and its ability to detect subtle, subclinical taste disorders. In our center, the test time of the EGM is 6 min, while the TST takes 12 min to test both sides of the tongue.

Over the years, several studies have been published on the validity, reliability, and diagnostic accuracy of the EGM.[3,4,6–22] Validity is the degree to which the EGM measures taste sensation. Reliability is the degree to which an electrical taste threshold is the same for patients who have not changed for repeated measurements under the same conditions.[23,24] Diagnostic accuracy is the ability of the EGM to distinguish between patients with and without taste loss.[25] To date, no systematic review of the validity, reliability, or diagnostic accuracy of the EGM has been published. To create an up-to-date clear overview of the published literature, we aim to systematically assess the validity, reliability, and diagnostic accuracy of the EGM in assessing taste sensation in adults.

2 | METHODS

2.1 | Search strategy and study selection

We conducted a systematic search in PubMed Medline, Elseviers's Embase, and the six databases of Cochrane Library with the assistance of a clinical librarian on December 20, 2022. We used a broad search query consisting of synonyms for EGM and reported our findings according to the PRISMA guideline.[26] The review has not been registered and no protocol has been prepared. See Appendix S1 for the full search strategy. No MeSH terms exist for EGM. No restrictions on publication year and publication status were applied. Two independent authors (EB and MD) excluded duplicates and performed title and abstract screening. See Figure 1 for the selection criteria. We considered randomized controlled trials and observational studies (retrospective and prospective cohort studies and case series [$n \geq 10$]) for inclusion if they included adults (≥18 years old) who underwent electrogustometry. Articles with original data were included for full text screening. Reviews, animal or laboratory studies, conference abstracts, poster presentations, letters, and case reports with <10 cases were excluded. Articles were also excluded if no analyses regarding validity, reliability, or diagnostic accuracy had been performed or if these analyses could not be performed with the published data. There were no restrictions regarding EGM model, stimulation method or study population. If no full text was available online or through our universities' libraries, the authors of the studies were contacted to retrieve the full text of the article. The full text screening was also performed by two independent authors (EB and DF) using the same selection criteria. Any differences were discussed until consensus was reached. Reference lists of included articles were used to identify possible additional relevant articles.

2.2 | Study outcome

The primary outcome was the construct validity of the EGM. Construct validity is the degree to which the scores of a measurement instrument are related to other measurement instruments that measure similar qualities. Therefore, the construct validity of the EGM is the degree to which the electrical taste threshold of the EGM is consistent with other taste instruments, such as TST and FPD.[24] The secondary outcomes were the reliability and diagnostic accuracy of the EGM. Reliability is the degree to which an electrical taste threshold is the same for patients who have not changed for repeated measurements under the same conditions.[23,24] Diagnostic accuracy is the ability of the EGM to distinguish between patients with and without taste loss.[25]

2.3 | Data extraction

Original data from included articles were extracted by two independent authors (EB and DF). The following data were extracted: country in which the study was performed, total number of participants, gender, age, study population, EGM model, stimulation method, stimulated nerve(s), duration of the stimulation, other taste instruments that were used and the measurement properties construct validity, reliability, and diagnostic accuracy. Construct validity and reliability are measured using correlation coefficients. A correlation coefficient between 0.1 and 0.3 was interpreted as weak, a correlation coefficient between 0.4 and 0.6 as moderate, a correlation between 0.7 and 0.9 as strong and a correlation of 1 as a perfect correlation.[27] Correlation coefficients were positive or negative. The lower the electrical taste threshold measured with the EGM, the better the taste sensation. If the reference test scale was the opposite of the EGM test scale (the higher the reference test score, the lower the electrical taste threshold), the correlation coefficient was negative. If the reference test scale was comparable to the EGM test scale (the lower the score of the reference test, the lower the electrical taste threshold), the correlation coefficient was positive. Diagnostic accuracy is measured using sensitivity, specificity, negative predictive value (NPV) and

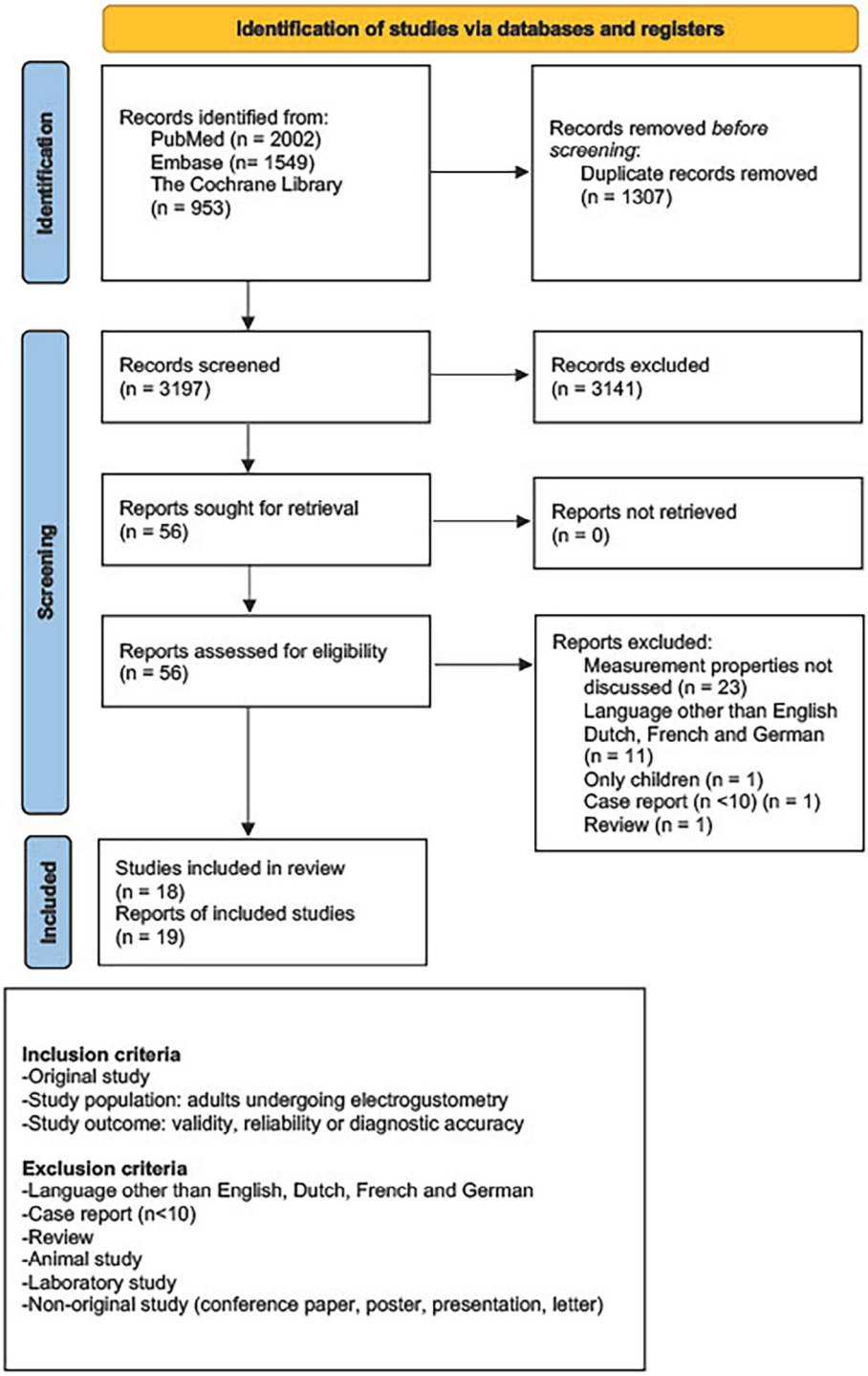

FIGURE 1 Prisma 2020 flow diagram.

positive predictive value (PPV). If data was not provided in the text of the article, it was extracted from figures if possible. If no measurement properties were reported, they were calculated.

2.4 | Quality assessment

Included studies were critically appraised by two independent authors (EB and DF). Predefined criteria from the COnsensus-based Standards for the selection of health status Measurements INstruments (COSMIN) tool were used to assess the quality of the studies on reliability (Appendix S2).[28,29] Each box was scored on a four-point rating scale (i.e., "very good," "adequate," "doubtful," or "inadequate"). The overall score, the quality of the study on reliability, was the lowest score given for a box. This method is called the worst-score-counts method.[28,29] The agreement on the scores was unanimous. The study designs and measurements used to quantify the validity and diagnostic accuracy of the EGM differed widely between the included studies. Therefore, we could not find one suitable tool to assess the quality of these studies. However, we assessed whether the studies formulated

hypothesis beforehand, had an appropriate sample size (>100 patients) to assess validity and whether there was any review bias.

3 | RESULTS

3.1 | Search and study selection

The electronic search of articles yielded 4504 results. After removing duplicates, 3197 articles were screened on title and abstract. Title and abstract screening resulted in 56 articles. During full text screening, 37 articles were excluded: one article was a case report with <10 cases[30]; one article only included children[31]; one article was a review[32]; the language of 11 articles was Japanese (five articles), Polish (four articles), or Russian (two articles)[33–43]; 23 articles did not discuss validity, reliability, or diagnostic accuracy of the EGM or did not report values that allowed us to calculate correlation coefficients, sensitivity, specificity, PPV, or NPV[44–66]; Nineteen articles were eligible for inclusion in the review.[3,4,6–22] Two articles from the same medical center and the same author evaluated the same patient cohort, so the results of these articles were combined.[14,15] We included the articles by Sakaguchi et al. and Saito et al., even though children were included in the study, because the mean age of the included patients was 39 and 42 years, respectively, and therefore we assumed only a small portion of the study population consisted of (young) children.[10,17]

3.2 | Risk of bias assessment

3.2.1 | Reliability studies

Six studies were included that assessed test–retest reliability of the EGM.[11,16–18,21,22] The majority of studies were not designed as a reliability study, which means that the methods used to assess test–retest reliability were poorly described (see Table 1).[11,17,18,21,22] Only the study by Nicolaescu et al. was designed as a reliability study.[16] Therefore, the overall quality of the included studies on reliability was rated as doubtful or inadequate, mostly due to the lack of information. All studies described the time interval between both tests, but none of the studies described if patients were clinically stable between the two measurements.[11,16–18,21,22] Nicolaescu et al. and Nilsson described the measurement protocol in detail for the first and the second measurement and the measurement conditions were similar for the repeated measurement.[16,22] The other studies did not describe their protocol in detail, but we assume measurement conditions were similar for the repeated measures.[11,17,18,21] None of the studies described whether professionals administered the measurements without knowledge of scores or values of the first measurement. Therefore, box four, administration of measurements, was rated as doubtful for all studies. Box five, assignment of the scores, does not apply to studies on reliability of the EGM, because no assignment of scores is necessary when using an EGM.[11,16–18,21,22] The electrical taste threshold is determined during the measurement. Berling et al. used ANOVA's and Nilsson used *t*-tests, which are important flaws in the statistical method, as these statistical methods are insufficient to test the reliability of the EGM.[11,22] The other four studies measured correlation coefficients, but none of the studies provided evidence that there was no systematic difference between measurements.[16–18,21] Only the study by Saito et al. calculated a Kappa, but the weighting scheme was not described.[17] Grant et al. did not report whether the correlation coefficient was statistically significant, which is also a major flaw in the statistical method.[21]

3.2.2 | Validity and diagnostic accuracy studies

Although we did not use a validated instrument to assess the risk of bias in the studies concerning validity and diagnostic accuracy, we think these studies carry a high risk of bias.[3,4,6–15,17–20] First of all, only the study by Ellegård et al. was designed as a validity and diagnostic accuracy study.[14,15] The other included studies retrospectively examined the validity and diagnostic accuracy of the EGM.[3,4,6–13,17–20] No hypotheses were therefore formulated in advance about the expected relationships or differences between the EGM and the reference tests. Second, it is unknown for all studies whether the EGM assessors were blinded and whether they had access to the results of the reference test. Also, the number of assessors was not mentioned in any of the studies.[3,4,6–15,17–20] Lastly, seven out of 12 studies on validity of the EGM did not have a sample size of more than 100 participants.[4,6,9,11,18–20]

3.3 | Characteristics of included studies

Table 2 shows the characteristics of the included studies. The studies were very heterogeneous with regard to study population,

TABLE 1 COSMIN risk of bias tool for reliability.

Reference (year)	1	2	3	4	5	6	7	8	9	Overall score
Berling (2011)[19]	D	V	A	D	NA	I	I	NA	NA	I
Nicolaescu (2005)[7]	D	V	V	D	NA	D	D	NA	NA	D
Saito (2001)[8]	D	V	A	D	NA	D	D	A	NA	D
Murphy (1995)[9]	D	V	A	D	NA	D	D	NA	NA	D
Grant (1987)[12]	D	V	A	D	NA	I	I	NA	NA	I
Nilsson (1977)[13]	D	V	V	D	NA	I	I	NA	NA	I

Abbreviations: V, very good; A, adequate; D, doubtful; I, inadequate; NA, not applicable.

TABLE 2 Characteristics of the included studies.

Reference (year)	Study design	Country	N (male %)	Age, years (mean [range])	Study population	EGM model	Stimulation method	Stimulated nerve(s)	Stimulation duration (ms)	Reference tests	Measurement properties
Pavlidis (2021)[2]	Prospective cohort study	Greece	132 (58)	NR	Healthy	TR-06, Rion Co., Tokyo, Japan	Staircase protocol	CT, N IX, major petrosal nerve	500	TST	Validity (construct)
Park (2021)[3]	Retrospective case series	South Korea	46 (24)	53 (NR)	Taste disorder with and without burning mouth symptoms	NR	Ascending protocol	CT, N IX, major petrosal nerve	NR	Subjective taste ability	Validity (construct)
Kang (2020)[14]	Prospective cohort study	South Korea	60 (50)	26 (NR)	Healthy	EGM-IIB, Nagashima Medical Instrument Co., Tokyo, Japan	Ascending protocol	CT, N IX, major petrosal nerve	NR	TST, FDP, taste solutions	Validity (construct)
Han (2018)[15]	Prospective case series	Germany	180 (37)	59 (NR)	Healthy, taste disorder, olfactory disorder	TR-06, Rion Co., Tokyo, Japan	Ascending protocol	CT	500	TST, taste solutions	Validity (construct)
Walliczek-Dworschak (2017)[16]	Prospective case series	Germany	121 (47)	52 (21–82)	Healthy, taste disorder	TR-06, Rion Co., Tokyo, Japan	Staircase protocol	CT	500	TST	Validity (construct)
Sakaguchi (2013)[18]	Prospective cohort study	Japan	126 (53)	39 (7–70)	Middle ear surgery	TR-06, Rion Co., Tokyo, Japan	NR	CT	500	Subjective taste ability	Diagnostic accuracy
Guder (2012)[17]	Prospective cohort study	Germany	18 (22)	45 (29–60)	Middle ear surgery	Halle II, Haberland, Halle, Germany	Ascending protocol	CT	500	Subjective taste ability	Validity (construct)
Berling (2011)[19]	Prospective cohort study	Sweden	39 (NR)	NR (27–62)	Healthy	TR-06, Rion Co., Tokyo, Japan	Ascending protocol	CT	2000	FPD	Validity (construct), Reliability (test–retest)
Deeb (2010)[20]	Prospective cohort study	United Kingdom	143 (NR)	NR	Early clinical signs of parkinsonism	TR-06, Rion Co., Tokyo, Japan	Staircase protocol	CT, N IX	1500	Subjective taste ability	Validity (construct)
Etoh (2008)[21]	Prospective case series	Japan	24 (71)	62 (NR)	Stroke with dysesthesia in the face, oral cavity or tongue (with or without taste disorder)	TR-06, Rion Co., Tokyo, Japan	Ascending protocol	CT	2000	FPD, subjective taste ability	Diagnostic accuracy

TABLE 2 (Continued)

Reference (year)	Study design	Country	N (male %)	Age, years (mean [range])	Study population	EGM model	Stimulation method	Stimulated nerve(s)	Stimulation duration (ms)	Reference tests	Measurement properties
Ellegård (2007)[5,6]	Prospective case series	Sweden	114 (42)	60 (NR)	Healthy, head and neck cancer treated with radiotherapy, Sjögren's syndrome, burning mouth syndrome or oral dysesthesia	Automated EGM, computer-controlled	Staircase protocol	CT	500	Subjective taste ability	Diagnostic accuracy, Validity (construct)
Nicolaescu (2005)[7]	Prospective cohort study	USA	16 (44)	24 (NR)	Healthy	TR-06, Rion Co., Tokyo, Japan	Staircase protocol	CT	500	NA	Reliability (test–retest)
Saito (2001)[8]	Retrospective cohort study	Japan	391 (NR)	42 (3–78)	Healthy, middle ear surgery	EGM-IIB, Nagashima Medical Instrument Co., Tokyo, Japan	Ascending protocol	CT	500	Subjective taste ability	Diagnostic accuracy, Reliability (test–retest)
Murphy (1995)[9]	Prospective cohort study	USA	17 (NR)	73 (64–85)	Healthy	TR-06, Rion Co., Tokyo, Japan	Ascending protocol	CT	500	Taste solutions	Validity (construct), Reliability (test–retest)
Grant (1989)[10]	Prospective case series	United Kingdom	20 (45)	43 (18–74)	Mastoidectomy, middle ear surgery and tympanoplasty	Programmable current source and a digital volt meter (RS components 258-041)	Ascending and descending protocol	CT	5000	Taste solutions	Validity (construct)
Le Floch (1989)[11]	Prospective case series	France	95 (48)	44 (NR)	Healthy, diabetes mellitus	NR	Staircase protocol	CT	NR	Taste solutions	Validity (construct)
Grant (1987)[12]	Prospective case series	United Kingdom	20 (NR)	NR	NR	Programmable current source and a digital volt meter (RS components 258-041)	Ascending protocol	CT	5000	NA	Reliability (test–retest)
Nilsson (1977)[13]	Prospective cohort study	Sweden	31 (61)	25 (NR)	Healthy	Madson GO 70 gustometer, Madsen Electronics, Copenhagen, Denmark	Ascending and descending protocol	CT	1000–1500	NA	Reliability (test–retest)

Abbreviations: *N*, sample size; EGM, electrogustometer; TST, taste strip test; FPD, filter paper disk; CT, chorda tympani; N IX, glossopharyngeal nerve; NR, not reported; NA, not applicable.

used measurements instruments and reported outcomes.[3,4,6–22] Major differences in type of patients and the reference tests that were used, precluded pooling of data.

3.3.1 | Study population

The mean age ranged from 24 to 73 years in 14 studies.[4,6–11,13–20,22] Only four studies included patients who underwent middle ear surgery.[9,10,17,19] The other studies included healthy participants and/or patients with a variety of diseases and syndromes.[3,4,6–8,11–20,22]

3.3.2 | Measurement instruments

The used reference tests were the TST, FDP, taste solutions, and subjective taste ability. Subjective taste ability was measured with a yes/no question,[10,13,17] visual analog scale (VAS),[9,14,15] or numeric rating scale (NRS).[4,12] The VAS used by Guder et al. ranged from 1 (no sense of taste) to 10 (excellent sense of taste).[9] The VAS used by Ellegård et al. ranged from 0 (no sense of taste) to 100 (excellent sense of taste).[14,15] Park et al. asked each patient to rate their sense of taste on a 3-point NRS from not at all recognized (0) to easily recognized (2).[4] Deeb et al. asked each patient to rate their sense of taste on a 6-point NRS from absent (1) to excellent (6).[12] The TR-06 (Rion Co., Tokyo, Japan) was the most commonly used EGM model.[3,7,8,10–13,16,18] Seven studies used a different EGM model and two studies did not report, which model was used.[4,6,9,14,15,17,19–22] Fifteen studies used the ascending protocol or the staircase detection protocol as the stimulation method.[3,4,6–9,11–18,20,21] Two studies used the ascending descending protocol and one study did not report the stimulation method.[10,19,22] During the ascending threshold protocol, an electrical stimulus in ascending microamps was applied. The rater started by applying the lowest electric current and increased it to the highest electric current. When patients identified the electrical taste, the specific current was considered to be the electrical taste threshold.[4,6,7,9,11,13,17] Park et al., Kang et al., and Han et al. performed the ascending threshold protocol three times and the mean value was considered to be the EGM threshold.[4,6,7] A two-alternative forced choice paradigm was used during the staircase detection protocol. Two correct answers or one incorrect answer led to a decrease or increase in stimulus intensity, which was a reversal in the staircase. The average of the staircase reversals was the electrical taste threshold.[3,8,12,14–16,20] Nilsson and Grant et al. used the descending protocol. After electric taste detection, the current was gradually reduced until the taste detection threshold was reached.[19,22] All studies stimulated the chorda tympani.[3,4,6–22] Pavlidis et al., Park et al., and Kang et al. also stimulated the glossopharyngeal nerve and the major petrosal nerve.[3,4,6] Deeb et al. also stimulated the glossopharyngeal nerve.[12] The chorda tympani carries the taste afferents of the anterior two thirds of the tongue. The glossopharyngeal nerve carries the taste afferents of the posterior one third of the tongue (circumvallate papillae). The major petrosal nerve carries the taste afferents of the soft palatal area.[3] Most studies stimulated for 500 ms.[3,7–10,14–18] Nilsson stimulated for 1000–1500 ms, Deeb et al. stimulated for 1500 ms, Berling et al. and Etoh et al. stimulated for 2000 ms and both studies by Grant et al. stimulated for 5000 ms.[11–13,19,21,22] Three studies did not report the duration of stimulation.[4,6,20]

3.4 | Primary outcome

Twelve studies evaluated construct validity by measuring correlation coefficients (Table 3). We reported the correlation coefficients between the EGM threshold in the chorda tympani area of the tongue and the reference test, because all studies stimulated this nerve.[3,4,6–9,11,12,15,18–20] Two studies did not report a correlation coefficient. They only described whether the EGM and the reference test were significantly correlated.[8,9] Ten studies showed a significant correlation between the EGM and a reference taste test. The level of statistical significance was set at a p-value of <0.05 for all studies.[3,4,6–8,11,12,15,19,20] There is no unequivocal answer to the question of which reference test correlates best with the EGM. Overall, the association between EGM and a reference taste test was weak or moderate (correlation coefficients ranged from −0.51 to 0.40).[3,4,6–9,11,12,15,18–20] Only Grant et al. found a strong correlation (correlation coefficient −0.74) between the EGM and taste solutions.[19] Three out of five studies using taste solutions as a reference test showed a significant correlation between the EGM and taste solutions,[7,19,20] whereas the other two did not find a significant correlation. Three out of four included studies using TST as a reference test showed a significant correlation between the EGM and TST. Pavlidis et al. and Walliczek-Dworschak et al. found a significant correlation between the EGM and all four taste strip scores (sweet score, salt score, sour score, and bitter score). However, Walliczek-Dworschak et al. did not report the values and they found negative correlations.[3,7,8] The correlations between the EGM and the FPD and between the EGM and subjective taste ability were generally weak with correlation coefficients around (−)0.3 and most of these were not statistically significant.[4,6,9,11,12,14]

Pavlidis et al. also found significant correlations, with similar values, between the electrical taste threshold of the glossopharyngeal nerve and the major petrosal nerve and the sweet, salty, sour and bitter scores of the TST. Park et al. found a significant correlation, with a similar value, between the electrical taste threshold of the major petrosal nerve and the sum score of the subjective taste ability in both the healthy and burning mouth study group. The electrical taste threshold of the major petrosal nerve was only associated with the sweet score in patients with burning mouth symptoms. Kang et al. and Deeb et al. found no significant correlations between the electrical taste threshold of the other nerves and the reference tests.[6,12]

3.5 | Secondary outcomes

3.5.1 | Test–retest reliability

Six studies evaluated test–retest reliability. The time between the first and the second measurement ranged from 1 day to 16 days.[11,16–18,21,22] Four studies calculated a correlation coefficient

TABLE 3 Results—Construct validity.

Test	Reference (year)	Correlations					
		Sweet score	Salt score	Sour score	Bitter score	Umami score	Sum score
TST	Pavlidis (2021)[2]	**0.40**	**0.40**	**0.25**	**0.40**	–	–
	Kang (2020)[14]	NC	NC	NC	NC	–	NC
	Han (2018)[15]						
	Healthy	NC	NC	NC	NC	NC	NC
	Taste disorder	**0.29**	NC	NC	NC	NC	NC
	Olfactory disorder	NC	NC	NC	NC	NC	NC
	Walliczek-Dworschak (2017)[16]	Negative correlation	Negative correlation	Negative correlation	Negative correlation	–	Negative correlation
FPD	Kang (2020)[14]	NC	NC	NC	NC	–	**0.38**
	Berling (2011)[19]	**0.29**	**0.27**	**0.31**	0.24	–	–
Taste solutions	Kang (2020)[14]	NC	NC	NC	NC	–	NC
	Han (2018)[15]						
	Healthy	NC	NC	**−0.34**	**−0.27**	NC	NC
	Taste disorder	0.29	NC	NC	NC	**−0.35**	**−0.38**
	Olfactory disorder	NC	NC	NC	NC	NC	NC
	Murphy (1995)[9]	–	−0.22	0.28	–	–	–
	Grant (1989)[10]	–	–	–	–	–	**−0.74**[a]
	Le Floch (1989)[11]	–	–	–	–	–	**−0.51**
Subjective taste ability	Park (2021)[3]						
	Burning mouth (right)	−0.38	−0.30	−0.39	−0.19	–	**−0.48**
	Burning mouth (left)	−0.15	−0.14	−0.27	0.00	–	−0.28
	No burning mouth (right)	−0.30	**−0.39**	−0.31	−0.25	–	−0.32
	No burning mouth (left)	−0.37	**−0.41**	**−0.40**	−0.39	–	**−0.44**
	Guder (2012)[17]	–	–	–	–	–	NC
	Deeb (2010)[20]	–	–	–	–	–	**−0.35**
	Ellegård (2007)[6]	−0.08	**−0.19**	−0.15	−0.12	–	–

Note: *p*-value <0.05 is statistically significant and displayed in bold.
Abbreviations: –, not reported; FPD, filter paper disk; NC, no correlation; TST, taste strip test.
[a]Calculated.

TABLE 4 Results—Diagnostic accuracy.

Reference (year)	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Sakaguchi (2013)[18]					
Subjective taste ability	33[a]	97[a]	68[a]	60[a]	98[a]
Etoh (2008)[21]					
FPD	75[a]	39[a]	83[a]	88[a]	32[a]
Subjective taste ability	46[a]	36[a]	69[a]	50[a]	56[a]
Ellegård. (2007)[5]					
Subjective taste ability		33 (4–71)	80 (71–87)	9 (1–28)	95 (88–99)
Sweet		0 (0–37)	77 (68–85)	0 (0–15)	91 (82–96)
Salt		40 (12–74)	81 (72–88)	17 (5–39)	93 (85–98)
Sour		30 (7–65)	80 (71–87)	13 (3–34)	92 (84–97)
Bitter					
Saito (2001)[8]					
Subjective taste ability	7[a]	80[a]	91[a]	40[a]	98[a]

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.
[a]Calculated.

between the first and second measurement. The first and second measurement were strongly associated (correlation coefficient ranged from 0.78 to 1.00).[16–18,21] The other two included studies did not report correlation coefficients, but found no significant threshold shift between the first and second measurement.[11,22] The level of statistical significance was also set at a *p*-value of <0.05 for all studies.[11,16–18,21,22]

3.5.2 | Diagnostic accuracy

Table 4 shows the prevalence, sensitivity, specificity, PPV, and NPV of the four included studies. The study population varied (see Table 2) and as a result the prevalence of taste loss ranged from 8% to 33%.[10,13,14,17] All studies used subjective taste ability as the reference test. Sensitivity ranged from 33% to 97%. Specificity was generally high, ranging between 68% and 91%. The PPV ranged from 0 to 88% with higher PPVs in study populations with a higher prevalence of taste disturbances. The NPVs were high in three studies with values ranging between 91% and 98%.[10,14,17] Etoh et al. reported lower NPVs, but also reported a higher prevalence of taste disturbances.

Etoh et al. also used the FPD as a reference test. Diagnostic measurement properties for the EGM were generally higher when the FPD was used as the reference test.

4 | DISCUSSION

4.1 | Summary of main results

The objective of this systematic review was to evaluate the validity, reliability, and the diagnostic accuracy of the EGM in assessing taste sensation. Nineteen articles discussing 18 studies were included for data extraction. The included studies on validity and diagnostic accuracy carry a high risk of bias and the overall quality of the studies on reliability were rated as doubtful or inadequate.[3,4,6–22] Overall, the association between a variety of reference taste tests and EGM was moderate or weak with correlation coefficients ranging from −0.51 to 0.40[3,4,6–9,11,12,15,18–20] with one outlier of −0.74 found in one study correlating EGM and taste solutions.[19] Test–retest reliability was good with reported correlation coefficients between 0.78 and 1.00.[16–18,21] The sensitivity, specificity, PPV, and NPV of EGM in identifying abnormal taste function varied widely between the four studies on diagnostic accuracy. Only the specificity was above 68% in all four studies.[10,13,14,17]

4.2 | Quality of evidence

The included studies on validity carry a high risk of bias. Eleven out of 12 included studies on validity retrospectively examined the validity of the EGM. Therefore, none of the included studies formulated hypothesis beforehand.[3,4,6–9,11,12,18–20] To investigate construct validity it is important to test predefined hypotheses, as it allows the authors to draw unbiased conclusions after data collection and analyses.[29] Furthermore, it is unknown for all studies whether the EGM assessors were blinded.[3,4,6–22] If the EGM assessors were not blinded, they were able to adjust their interpretation of the electric threshold to match the results of the reference test, which is called review bias. Review bias can falsely increase validity and diagnostic accuracy of the EGM.[67] Finally, a sample size of more than 100 patients is an appropriate number to test validity and reliability, but only seven of 14 studies evaluating validity and/or reliability had a sample size of more than 100 participants.[24]

4.3 | Factors that influence electrical taste threshold

Older age, smoking, a longer duration of electrical stimuli, a smaller electrode and stimulation of a tongue region less densely populated by taste buds are associated with a higher electrical taste threshold. Electrical taste thresholds are higher in patients older than 60 years compared to younger patients and in smokers compared to non-smokers.[58] Pavlidis et al. increased the stimulus duration from 500 to 2000 ms and found that the electrical taste threshold increased with longer stimulus duration.[3] Nicolaescu et al. found that a 25 mm^2 electrode is associated with a higher electrical taste threshold compared to a 125 mm^2 electrode.[16] Their hypothesis is that a smaller electrode reflects the stimulation of fewer taste buds. Also, a larger electrode results in more reliable threshold values.[16,68] Taste thresholds are higher in tongue regions that are less densely populated by taste buds, such as the medial tongue region, than tongue regions that are more densely populated, such as the tongue tip.[16]

The method of detecting the electrical taste threshold varies. Most studies used the ascending protocol or the staircase detection protocol.[3,4,6–9,11–19] None of the included studies compared different stimulation methods.[3,4,6–22] Staircase procedures have been widely used in measuring thresholds of sensory systems. Therefore, this is currently the recommended stimulation method when using the EGM.[69,70] The recommended procedure during the staircase detection protocol is to continue testing until at least six reversals are obtained.[69,70] Ellegård et al. obtained eight reversals per patient, following the protocol of Stillman et al., and Nicolaescu et al. obtained seven reversals per patients.[14–16,52] The other four studies did not document the number of reversals or obtained fewer than six reversals.[3,8,12,20]

Nine out of the 18 included studies reported that the ion current during electrical stimulation was anode (+).[10,14–22] The other nine studies did not report whether the anode or cathode (−) touched the tongue during EGM testing.[3,4,6–9,11–13] Anodic stimulation results in a sour taste, while cathodic stimulation leads to a bitter/soapy taste. The advantage of using the anode is that the risk of trigeminal stimulation being perceived by the subject while measuring electrical taste thresholds is lower. Trigeminal stimulation is sensed when using higher current levels, causing a tingling or burning stimulation.[3,51,63] Cathodic stimulation only evokes taste at higher currents, so using a cathode ion current leads to higher risk of trigeminal nerve stimulation perceived by the subjects when measuring electrical taste

thresholds.[71] To avoid trigeminal involvement, participants should be instructed to distinguish between the sensation experienced by electrical stimulation of the gustatory nerve and the sensation experienced by electrical stimulation of the trigeminal nerve.[3]

4.4 | Strengths and weaknesses of review process

To our knowledge, this is the first systematic review on this topic. The search strategy was transparent and developed with the help of a clinical librarian. The title abstract screening, full text screening, quality assessment, and data extraction of eligible articles was performed independently by two authors. The evaluation of validity, reliability, and diagnostic accuracy was based on predetermined criteria, which were described in the methods section.

Several limitations may have biased the outcomes of our systematic review. First of all, we did not use a risk of bias tool to assess the quality of the studies on validity and diagnostic accuracy. We evaluated the quality of the includes studies, but a risk of bias tool helps to transparently evaluate the quality of the included studies by the authors. However, we would have had to use numerous different tools to assess the risk of bias in these studies, because the study design and measurements used to quantify the diagnostic accuracy or validity of the EGM differed widely between the studies. Second, study populations and experimental procedures differed between studies. As a result, data could not be pooled. Lastly, articles written in languages other than English, Dutch, French, and German were excluded during title abstract and full text screening. Therefore, we excluded articles written in Japanese, Russian, and Polish. There is a possibility that these articles were relevant to our review.

5 | CONCLUSION

The included studies in this review lack the required standard of the study design to draw firm conclusions about the validity, reliability and diagnostic accuracy of the EGM. Future research is needed to assess the validity, reliability, and diagnostic accuracy of the EGM. However, based on the reported results, we would not recommend using the EGM as a screening test for taste disturbance in clinical practice.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data, code, and other materials will be available upon reasonable request.

ORCID

Max J. Dullaart https://orcid.org/0000-0002-4435-7007
Henricus G. X. M. Thomeer https://orcid.org/0000-0003-0937-6189
Inge Wegner https://orcid.org/0000-0002-4124-7224

REFERENCES

1. Krarup B. Electro-gustometry: a method for clinical taste examinations. *Acta Otolaryngol*. 1958;49(4):294-305.
2. Stillman JA, Morton RP, Hay KD, Ahmad Z, Goldsmith D. Electrogustometry: strengths, weaknesses, and clinical evidence of stimulus boundaries. *Clin Otolaryngol Allied Sci*. 2003;28(5):406-410.
3. Pavlidis P, Schittek GA, Saratziotis A, Ferfeli M, Kekes G, Gouveris H. Electrogustometry: normative data for stimulus duration, tongue site and age decline. *Clin Otolaryngol*. 2021;46(4):767-774.
4. Park YJ, Kim MJ, Kho HS. Relationships between subjective taste sensations and electrogustometry findings in patients with taste disorders. *Int J Oral Maxillofac Surg*. 2021;50(4):522-529.
5. Mueller C, Kallert S, Renner B, et al. Quantitative assessment of gustatory function in a clinical context using impregnated "taste strips". *Rhinology*. 2003;41(1):2-6.
6. Kang MG, Choi JH, Kho HS. Relationships between gustatory function tests. *Oral Dis*. 2020;26(4):830-837.
7. Han P, Georgi M, Cuevas M, Haehner A, Gudziol V, Hummel T. Decreased electrogustometric taste sensitivity in patients with acquired olfactory dysfunction. *Rhinology*. 2018;56(2):158-165.
8. Walliczek-Dworschak U, Schöps F, Feron G, Brignot H, Hähner A, Hummel T. Differences in the density of fungiform papillae and composition of saliva in patients with taste disorders compared to healthy controls. *Chem Senses*. 2017;42(8):699-708.
9. Guder E, Böttcher A, Pau HW, Just T. Taste function after stapes surgery. *Auris Nasus Larynx*. 2012;39(6):562-566.
10. Sakaguchi A, Nin T, Katsura H, Mishiro Y, Sakagami M. Trigeminal and taste sensations of the tongue after middle ear surgery. *Otol Neurotol*. 2013;34(9):1688-1693.
11. Berling K, Knutsson J, Rosenblad A, Von Unge M. Evaluation of electrogustometry and the filter paper disc method for taste assessment. *Acta Otolaryngol*. 2011;131(5):488-493.
12. Deeb J, Shah M, Muhammed N, et al. A basic smell test is as sensitive as a dopamine transporter scan: comparison of olfaction, taste and DaTSCAN in the diagnosis of Parkinson's disease. *QJM*. 2010;103(12):941-952.
13. Etoh S, Kawahira K, Ogata A, Shimodozono M, Tanaka N. Relationship between dysgeusia and dysesthesia in stroke patients. *Int J Neurosci*. 2008;118(1):137-147.
14. Ellegård EK, Hay KD, Morton RP. Is electrogustometry useful for screening abnormalities of taste? *J Laryngol Otol*. 2007;121(12):1161-1164.
15. Ellegård EK, Goldsmith D, Hay KD, Stillman JA, Morton RP. Studies on the relationship between electrogustometry and sour taste perception. *Auris Nasus Larynx*. 2007;34(4):477-480.
16. Nicolaescu SA, Wertheimer JM, Barbash SE, Doty RL. Electrical taste thresholds established on the medial tongue using two sizes of electrodes. *Laryngoscope*. 2005;115(8):1509-1511.
17. Saito T, Manabe Y, Shibamori Y, et al. Long-term follow-up results of electrogustometry and subjective taste disorder after middle ear surgery. *Laryngoscope*. 2001;111(11 Pt 1):2064-2070.
18. Murphy C, Quiñonez C, Nordin S. Reliability and validity of electrogustometry and its application to young and elderly persons. *Chem Senses*. 1995;20(5):499-503.
19. Grant R, Miller S, Simpson D, Lamey PJ, Bone I. The effect of chorda tympani section on ipsilateral and contralateral salivary secretion and taste in man. *J Neurol Neurosurg Psychiatry*. 1989;52(9):1058-1062.
20. Le Floch JP, Le Lievre G, Sadoun J, Perlemuter L, Peynegre R, Hazard J. Taste impairment and related factors in type I diabetes mellitus. *Diabetes Care*. 1989;12(3):173-178.
21. Grant R, Ferguson MM, Strang R, Turner JW, Bone I. Evoked taste thresholds in a normal population and the application of electrogustometry to trigeminal nerve disease. *J Neurol Neurosurg Psychiatry*. 1987;50(1):12-21.

22. Nilsson B. Taste acuity of the human palate I. studies with electrogustometry and taste solutions on young adults. *Acta Odontol Scand*. 1977;35(1-3):51-62.
23. Bannigan K, Watson R. Reliability and validity in a nutshell. *J Clin Nurs*. 2009;18(23):3237-3243.
24. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539-549.
25. Šimundić AM. Measures of diagnostic accuracy: basic definitions. *EJIFCC*. 2009;19(4):203-211.
26. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
27. Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med*. 2018;18(3):91-93.
28. Mokkink LB, Boers M, van der Vleuten CPM, et al. COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med Res Methodol*. 2020;20(1):1-18.
29. Mokkink LB, Prinsen CAC, Patrick DL, et al. *COSMIN Study Design checklist for Patient-reported outcome measurement instruments—user manual 2019*. https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf; 2019:1-31.
30. Lobb B, Elliffe DM, Stillman JA. Reliability of electrogustometry for the estimation of taste thresholds. *Clin Otolaryngol Allied Sci*. 2000;25(6):531-534.
31. Bennetto L, Kuschner ES, Hyman SL. Olfaction and taste processing in autism. *Biol Psychiatry*. 2007;62(9):1015-1021.
32. Tomita H, Ikeda M, Okuda Y. Basis and practice of clinical taste examinations. *Auris Nasus Larynx*. 1986;13:S1-S15.
33. Yaginuma Y, Kobayashi T, Sai Y, Takasaka T. Predictive value of electrogustometry in the preoperative diagnosis of severity of middle ear pathology. *Nihon Jibiinkoka Gakkai Kaiho*. 1996;99(11):1635-1640.
34. Sawada M. A study of measurements of and factors influencing threshold levels of taste perception. *Kokubyo Gakkai Zasshi the Journal of the Stomatological Society, Japan*. 2005;72(1):28-41.
35. Nin T, Umemoto M, Miuchi S, Negoro A, Sakagami M. Treatment outcome in patients with taste disturbance. *Nihon Jibiinkoka Gakkai Kaiho*. 2006;109(5):440-446.
36. Nakamura H, Azuma H, Kawamoto M, Ito A, Isono M, Murata K. Reliability of electrogustatory threshold—electrogustometry using CNV. *Nihon Jibiinkoka Gakkai Kaiho*. 2000;103(10):1161-1168.
37. Kuga M, Ikeda M. Evaluation of gustatory threshold changes in healthy subjects. *Nihon Jibiinkoka Gakkai Kaiho*. 1996;99(3):411-416.
38. Jatczak J, Kordasz P. Repetition time of taste perception in electrogustometric tests. *Czas Stomatol*. 1978;31(10):907-909.
39. Jatczak J, Kordasz P. Taste perception threshold deviation during electrogustometric measurements. *Protet Stomatol*. 1978;28(1):15-19.
40. Blagoveshchenskaia NS, Mukhamedzhanov NZ. Taste studied by electrogustometry in otorhinolaryngology. *Vestn Otorinolaringol*. 1980;1:27-31.
41. Blagoveshchenskaia NS, Mukhamedzhanov NZ. Electrometric method of studying taste in brain diseases. *Zhurnal Voprosy Neirokhirurgii Imeni N N Burdenko*. 1980;3:47-52.
42. Kordasz P. The usefulness of electrogustometer for taste examination. *Protet Stomatol*. 1976;26(1):23-27.
43. Dobosz P, Konopka W, Grzanka A. An evaluation of the usefulness of impulse-current stimulation in electrogustometry—preliminary report. *Pol Merkur Lekarski*. 2005;19(111):294-295.
44. Al-Ezzi M, Khan K, Tappuni AR. Is the taste acuity affected by oral dryness in primary Sjögren's syndrome patients? *Oral Dis*. 2020;26(3):688-695.
45. Kuba S, Fujiyama R, Yamanouchi K, et al. Awareness of dysgeusia and gustatory tests in patients undergoing chemotherapy for breast cancer. *Support Care Cancer*. 2018;26(11):3883-3889.
46. Maeda E, Katsura H, Nin T, Sakaguchi-Fukunaga A, Mishiro Y, Sakagami M. Change of somatosensory function of the tongue caused by chorda tympani nerve disorder after stapes surgery. *Laryngoscope*. 2018;128(3):701-706.
47. Doty RL, Heidt JM, MacGillivray MR, et al. Influences of age, tongue region, and chorda tympani nerve sectioning on signal detection measures of lingual taste sensitivity. *Physiol Behav*. 2016;155:202-207.
48. Saito T, Ito T, Narita N, Yamada T, Manabe Y. Light and electron microscopic observation of regenerated fungiform taste buds in patients with recovered taste function after severing chorda tympani nerve. *Ann Otol Rhinol Laryngol*. 2011;120(11):713-721.
49. Stevens DA, Baker D, Cutroni E, Frey A, Pugh D, Lawless HT. A direct comparison of the taste of electrical and chemical stimuli. *Chem Senses*. 2008;33(5):405-413.
50. Kitagoh H, Tomita H, Ikui A, Ikeda M. Course of recovery from taste receptor disturbance. *Acta Otolaryngol Suppl*. 2002;6489(546):83-93.
51. Tomita H, Ikeda M. Clinical use of electrogustometry: strengths and limitations. *Acta Otolaryngol Suppl*. 2002;6489(546):27-38.
52. Stillman JA, Morton RP, Goldsmith D. Automated electrogustometry: a new paradigm for the estimation of taste detection thresholds. *Clin Otolaryngol Allied Sci*. 2000;25(2):120-125.
53. Arnold SM. The vulnerability of the chorda tympani nerve to middle ear disease. *J Laryngol Otol*. 1974;88(5):457-466.
54. Fons M, Aabo OP. Electro-Gustometry part 1 intensity relations of electrical taste. *J Laryngol Otol*. 1980;224(1348):1073-1075.
55. Norré ME, Suykerbuyk L. Critical study of EGM as clinical test. *Acta Otorhinolaryngol Belg*. 1979;33(6):936-943.
56. Just T, Homoth J, Graumüller S, Pau HW. Taste disorders and recovery of the taste function after middle ear surgery. *Laryngorhinootologie*. 2003;82(7):494-500.
57. Just T, Pau HW, Bombor I, Guthoff RF, Fietkau R, Hummel T. Confocal microscopy of the peripheral gustatory system: comparison between healthy subjects and patients suffering from taste disorders during radiochemotherapy. *Laryngoscope*. 2005;115(12):2178-2182.
58. Pavlidis P, Gouveris H, Anogeianaki A, Koutsonikolas D, Anogianakis G, Kekes G. Age-related changes in electrogustometry thresholds, tongue tip vascularization, density, and form of the fungiform papillae in humans. *Chem Senses*. 2013;38(1):35-43.
59. Sienkiewicz-Jarosz H, Scinska A, Kuran W, et al. Taste responses in patients with Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 2005;76(1):40-46.
60. Coats AC. Normal limit of the electrogustometry test. *Ann Otol Rhinol Laryngol* 1974;83(4):491-497.
61. Kikuchi T, Kusakari J, Kawase T, Takasaka T. Electrogustometry of the soft palate as a topographic diagnostic method for facial paralysis. *Acta Otolaryngol Suppl*. 1988;458:134-138.
62. Negoro A, Umemoto M, Fujii M, et al. Taste function in Sjögren's syndrome patients with special reference to clinical tests. *Auris Nasus Larynx*. 2004;31(2):141-147.
63. Miller SL, Mirza N, Doty RL. Electrogustometric thresholds: relationship to anterior tongue locus, area of stimulation, and number of fungiform papillae. *Physiol Behav*. 2002;75(5):753-757.
64. Loucks CA, Doty RL. Effects of stimulation duration on electrogustometric thresholds. *Physiol Behav*. 2004;81(1):1-4.
65. Pavlos P, Vasilios N, Antonia A, Dimitrios K, Georgios K, Georgios A. Evaluation of young smokers and non-smokers with Electrogustometry and contact endoscopy. *BMC Ear Nose Throat Disord*. 2009;9:9.
66. Braud A, Descroix V, Ungeheuer MN, Rougeot C, Boucher Y. Taste function assessed by electrogustometry in burning mouth syndrome: a case-control study. *Oral Dis*. 2017;23(3):395-402.

67. Kea B, Hall MK, Wang R. Recognising bias in studies of diagnostic tests part 2: interpreting and verifying the index test. *Emerg Med J*. 2019;36(8):501-505.
68. Doty RL. Measurement of chemosensory function. *World J Otorhinolaryngol Head Neck Surg*. 2018;4(1):11-28.
69. Wetherill GB, Levitt H. Sequential estimation of points on a psychometric function. *Br J Math Stat Psychol*. 1965;18:1-10.
70. Levitt H. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*. 1971;49(2):467+.
71. Hettinger TP, Frank ME. Salt taste inhibition by cathodal current. *Brain Res Bull*. 2009;80(3):107-115.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Blijleven EE, Fuchten D, Dullaart MJ, Stokroos RJ, Thomeer HGXM, Wegner I. Systematic review: Validity, reliability, and diagnostic accuracy of the electrogustometer. *Laryngoscope Investigative Otolaryngology*. 2023;8(4):1068-1079. doi:10.1002/lio2.1108