Received: 17 October 2019 | Revised: 8 January 2020 | Accepted: 11 February 2020

DOI: 10.1002/hbm.24957

RESEARCH ARTICLE

WILEY

Empirical assessment of changing sample-characteristics in task-fMRI over two decades: An example from gustatory and food studies

Andy W. K. Yeung[1] | **Natalie S. M. Wong**[2] | **Simon B. Eickhoff**[3,4]

[1]Oral and Maxillofacial Radiology, Applied Oral Sciences and Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China

[2]Oral and Maxillofacial Surgery, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China

[3]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[4]Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Centre Jülich, Jülich, Germany

Correspondence
Andy W. K. Yeung, 1B39B, Prince Philip Dental Hospital, 34 Hospital Road, Hong Kong, China.
Email: ndyeung@hku.hk

Simon B. Eickhoff, Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
Email: Simon.Eickhoff@uni-duesseldorf.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: EI 816/11-1; Horizon 2020 Framework Programme, Grant/Award Number: 785907 (HBP SGA2); National Institute of Mental Health, Grant/Award Number: R01-MH074457; Projekt Deal, Grant/Award Number: EI 816/11-1

Abstract

Over the past two decades, functional neuroimaging has not only grown into a large field of research, but also substantially evolved. Here we provide a quantitative assessment of these presumed in sample composition and data analysis, using fMRI studies on food/taste research published between 1998 and 2019 as an exemplary case in which the scientific objectives themselves have remained largely stable. A systematic search for papers written in English was done using multiple databases and identified 426 original articles that were subsequently analyzed. The median sample size significantly increased from 11.5 to 35.5 while the ratio of male to female subjects remained stable. There were, however, more papers involving female subjects only, rather than male subjects only, since 2003. There was a decline in uncorrected results and statistical correction by false-discovery rate. Reflecting a trend toward more conservative thresholding, the number of foci reported per paper did not change significantly and sample size (power) did not correlate with the number of reported foci. The median journal impact factor and the normalized number of citations (citations per year) of the papers, in turn, showed a significantly decreasing trend. Number of citations negatively correlated to sample size, publication year but positively correlated to journal impact factor, and was also influenced by statistical correction method. There was a decreasing trend in studies recruiting both left-handed and right-handed subjects. In summary, the present paper quantifies several large-scale trends that have often been anecdotally discussed and reveals the changing nature of neuroimaging studies that may be considered when pursuing meta-analytic approaches.

KEYWORDS

bias, fMRI, food, gustation, handedness, neuroimaging, sex, taste

1 | INTRODUCTION

The neuroimaging literature contains many small sample studies that could be underpowered and with modest replicability (Button et al., 2013; Turner, Paul, Miller, & Barbey, 2018). The small samples created an issue of representativeness. Can results generated by those small samples be generalized to other populations and become relevant to us? A recent study surveyed approximately 1,000 neuroimaging articles, only to find that 3–4% of recruited subjects were left-handed or ambidextrous (Bailey, McMillan, & Newman, 2019). In neuroscience, nearly half of the animal studies used males only, with 25% used both sexes and 9% used female only (Beery, 2018). When both human and animal studies were considered, there were increasing numbers of articles reported sex from the beginning of the 2010s to the mid-2010s, but sex bias remained present, with more and more articles involved only males (Will et al., 2017). It is worthwhile to revisit the human functional magnetic resonance imaging (fMRI) literature to see if these limitations have improved over time, and if they influenced the citation count and other measurable outcomes, such as number of reported activated foci.

Moreover, studies with larger samples should have more power to detect significant findings, such as activated voxels or clusters, and therefore have a larger number of reported foci. This notion was tested in fMRI studies in general, voxel-based morphometry studies of psychiatric disorders, and neuroimaging studies of sex differences (David et al., 2013; David et al., 2018; Fusar-Poli et al., 2014). Results found that these bodies of literature have potential reporting bias, as the correlation between the two parameters (sample size and number of foci) was either insignificant or very weak (David et al., 2013; David et al., 2018; Fusar-Poli et al., 2014). Meanwhile, it was unknown if the underpowered studies with small samples were cited more frequently than their counterparts with larger samples, thus creating a citation bias and increasing the risk of disseminating some findings potentially to be false positives.

The evolution of neuroimaging samples was still largely unknown. In food and taste neuroscience, there was a guideline that recommended items and details to be reported (Smeets et al., 2019), multiple meta-analyses to answer specific research questions (Huerta, Sarkar, Duong, Laird, & Fox, 2014; Veldhuizen et al., 2011; Yeung, Goto, & Leung, 2017, 2018), and even a meta-evaluation to evaluate the reporting quality of these meta-analyses (Yeung, Wong, Lau, & Eickhoff, 2019). Being a moderately sized literature, big enough to be meaningful but small enough for us to handle manually, the food and taste fMRI literature is a perfect example for us to conduct a case-study on the evolution of the research landscape. Did the literature evolve over time in terms of subject composition and data analytics? We addressed this by surveying a body of literature mainly published over the last two decades, to reveal any obvious trends across time.

2 | MATERIALS AND METHODS

2.1 | Search strategy

We searched three literature databases in September 2019: BrainMap (http://apps.rii.uthscsa.edu/bmapWeb/), Neurosynth (https://neurosynth.org/), and PubMed (https://www.ncbi.nlm.nih.gov/pubmed/). For BrainMap, we searched twice, once with filters of “stimulus modality = gustatory” and “imaging modality = fMRI”; and once with filters of “stimulus type = food” and “imaging modality = fMRI.” For Neurosynth, we clicked “meta-analysis,” chose “taste” and “food” from the terms list, and manually excluded the non-fMRI papers. For PubMed, we searched for (“functional magnetic resonance imaging” OR “MRI” OR “BOLD”) AND (“taste” OR “gustatory” OR

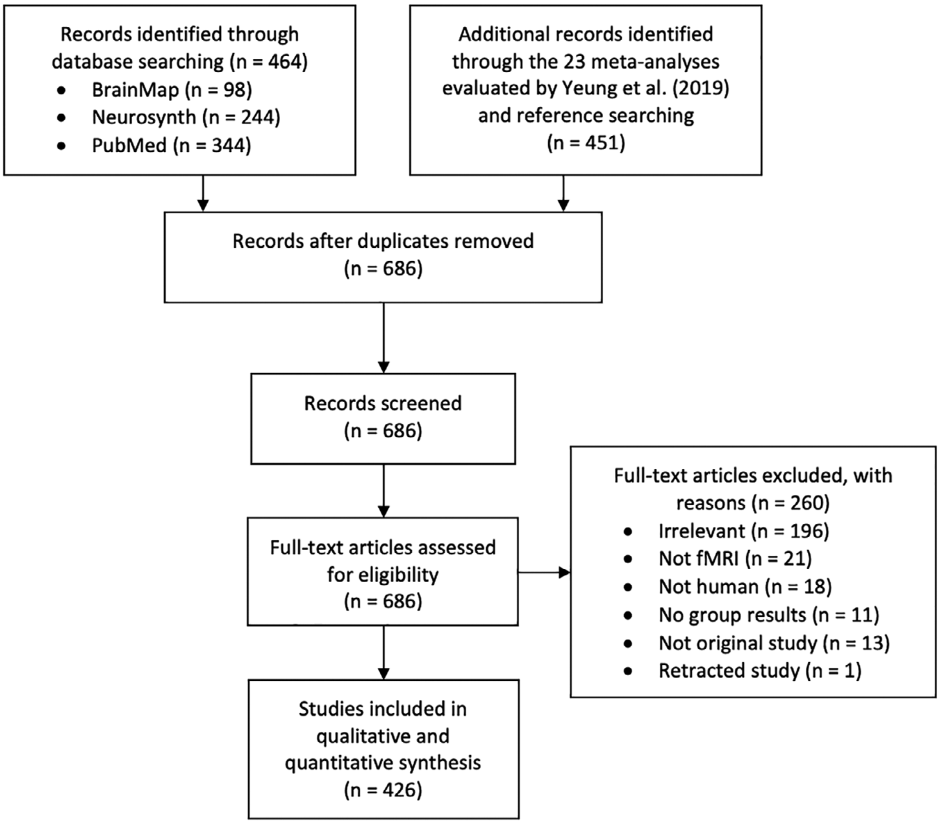


FIGURE 1 PRISMA diagram of the study workflow

"gustation" OR "tastants" OR "flavor") in the title/abstract fields. Publications were restricted to be those written in English. No restriction was placed on the publication year. Full texts of the publications were assessed to exclude the ineligible ones. The exclusion criteria were irrelevance, non-fMRI study, non-human study, absence of group results (i.e., single subject studies, or studies reporting results from individual but not group analyses), non-original study, no reporting of brain coordinates, and retracted study. At the end of the screening process, 426 articles were included into the analysis. Figure 1 shows the PRISMA diagram for the screening process.

2.2 | Data extraction

For each of the 426 included article, the following parameters were extracted: total number of reported foci; total sample size; publication year; ratio of male subjects; ratio of right-handed subjects; subject type (healthy controls or patients); health problems focused; stimuli type (e.g., tastant, visual, or mixed); nature of investigation (e.g., somatosensory, hedonics, cognition, or mixed); magnet strength; identity of fMRI data processing software; involvement in whole-brain analysis (whole-brain analysis was involved, or only region-of-interest (ROI) analysis was performed); strictest statistical correction employed (familywise error [FWE], false discovery rate [FDR], unclear, uncorrected, or unknown); assessment of task-based activations, resting state brain connectivity, functional connectivity, and effective connectivity; 2018 Journal Impact Factor; and total citation count. Two authors (AY and NW) did the data extraction independently, and any disagreements were resolved by discussion and reaching a consensus. During data extraction, the number of foci discounted the ones irrelevant to food and taste (e.g., if a study investigated reward using tastant and photos of money, the foci resulted from contrasts solely dealing with money would be excluded). For articles that recruited a uni-sex sample, or a sample that was totally right-handed (or totally left-handed), we searched if the reason was explicitly stated. The common correlational analyses between brain activity level and behavioral score were recorded as brain mapping. Meanwhile, psychophysiological interaction (PPI) was recorded as an involvement in functional connectivity but not effective connectivity (O'Reilly, Woolrich, Behrens, Smith, & Johansen-Berg, 2012). The number of citations and 2018 Journal Impact Factor were extracted from the Web of Science Core Collection electronic database hosted by Clarivate Analytics (https://www.webofknowledge.com/).

2.3 | Statistical analysis

The time trends of extracted parameters were evaluated. The major outcome variables namely the sample size, the number of citations, and the number of foci of the articles were not normally distributed (Shapiro–Wilk test: $W = 0.467$, 0.704, and 0.599, respectively, $p < .001$), and the relationship between them were evaluated by Spearman correlation. The number of citations and the number of foci were tested to see if there were significant influencing factors among the other extracted parameters, such as publication year, male subject ratio, right-handed subject ratio, 2018 Journal Impact Factor, subject type, healthy problem focused, stimuli type, investigation nature, magnet strength, fMRI software, whole brain analysis involvement (vs. only ROI analysis), statistical correction, connectivity analysis involvement (vs. task-based activation only). Two-tailed Spearman correlation, Kruskal-Wallis test, and Mann Whitney test were used whenever appropriate. Statistical tests were performed with SPSS 25.0 (IBM, NY), and results were significant if $p < .05$.

3 | RESULTS

3.1 | Cross-sectional results

The 426 analyzed articles were published between 1998 and 2019, and had a median sample of 24 subjects and a median number of foci reported of 17.5 (Table 1). The median male subject ratio was 0.46, and median right-handed subject ratio was 1.00. The median citation count of the articles was 32.5. The raw data of the articles can be found in Supporting Information.

Around 3.8% of the 426 articles ($n = 16$) did not mention the sex of the subjects, whereas 12.4% ($n = 53$) recruited both sexes in exact equal proportion. About one-fourth ($n = 106$) of the 426 articles recruited female subjects only, and 7.3% ($n = 31$) recruited males only (Figure 2a). Among the 106 all females papers, 35.8% were non-clinical (not investigating any health problems), whereas the remaining ones were mostly investigating eating disorder (31.1%) and obesity related issues (29.2%). On the contrary, 74.2% of the all males' papers were non-clinical, whereas the rest were mainly investigating drinking related issues (16.1%). Over 70% of these 137 uni-sex articles ($n = 96$) did not explicitly elaborate on this decision. For the remaining 41 articles that stated a reason, the common reasons were to avoid confounding effects of sex ($n = 14$), investigate diseases predominant or specific in females (e.g., eating disorders and polycystic ovary syndrome; $n = 12$), avoid the variability/effects of menstrual cycle ($n = 5$), females are more reactive ($n = 4$), investigate sex- or population-specific effects ($n = 3$), males are more reactive ($n = 1$), have difficulty in recruiting the other sex due to complex eligibility criteria ($n = 1$), and merely following a previous study ($n = 1$).

Over one-third of the 426 articles ($n = 159$) did not mention the handedness of the subjects. For the remaining 267 articles, nearly 90% ($n = 238$) recruited all right-handed subjects (Figure 2b), whereas only 0.5% ($n = 2$) recruited left-handed and right-handed subjects in equal proportion. Only two of the all right-handed articles justified the decision by stating that it wanted to avoid confounds caused by handedness. Besides, no article recruited mainly left-handed subjects.

Nearly half of the studies (48.9%) recruited healthy subjects only (Table 1). Meanwhile, the most commonly investigated health problems from the literature were:

TABLE 1 The number of foci reported in the main text and the number of citations of the food and taste fMRI papers

Overall parameters	Median (IQR)	Mean (SD)	Range
Sample size	24.0 (22)	33.9 (41.7)	3–416
No. of foci reported in main text	17.5 (25)	25.7 (29.7)	0–249
Publication year	2012 (5)	2011.9 (4.5)	1998–2019
Male subject ratio	0.46 (0.55)	0.39 (0.30)	0–1.0
Right-handed subject ratio	1.00 (0.00)	0.98 (0.06)	0.50–1.00
No. of Web of Science citations	32.5 (75)	73.4 (118.1)	0–930
2018 Journal Impact Factor	3.749 (3.028)	5.096 (4.498)	0–41.037

Specific factors	Sample size, median (IQR)	No. of foci, median (IQR)	No. of citations, median (IQR)
1. Subject type			
Healthy (*n* = 208)	20.0 (16)	18.0 (24)	30.0 (73)
Patients (*n* = 57)	26.0 (24)	16.0 (34)	28.0 (39)
Mixed (*n* = 156)	33.0 (23)	18.0 (26)	42.0 (90)
Unknown (*n* = 5)	24.0 (43)	8.0 (10)	122.0 (410)
2. Healthy problem focused			
None (*n* = 218)	20.0 (17)	17.5 (24)	32.0 (77)
Eating disorder-related (*n* = 35)	39.0 (24)	22.0 (22)	34.0 (84)
Obesity-related (*n* = 91)	30.0 (22)	18.0 (26)	34.0 (65)
Drinking-related (*n* = 52)	31.5 (44)	15.0 (27)	41.0 (110)
Syndrome-related (*n* = 11)	18.5 (16)	24.0 (35)	10.0 (49)
Psychiatric-related (*n* = 13)	33.0 (23)	18.0 (12)	17.0 (53)
Others (*n* = 6)	22.0 (17)	8.0 (7)	11.5 (25)
3. Treatment/drug evaluation			
No (*n* = 385)	24.0 (22)	18.0 (26)	33.0 (79)
Yes (*n* = 41)	27.0 (21)	15.0 (16)	30.0 (51)
4. Stimuli type			
Visual (*n* = 215)	26.0 (22)	18.0 (25)	39.0 (83)
Tastant (*n* = 131)	21.0 (21)	17.0 (26)	19.0 (52)
Odor (*n* = 8)	19.0 (15)	13.5 (34)	40.0 (68)
Auditory (*n* = 1)	9	9	56
Electric taste (*n* = 1)	11	91	38
Mixed (*n* = 70)	26.5 (23)	18.0 (23)	33.5 (110)
5. Investigation nature			
Somatosensory (*n* = 160)	20.0 (23)	16.5 (26)	32.0 (62)
Hedonics (*n* = 183)	26.0 (24)	18.0 (23)	32.0 (87)
Cognition (*n* = 57)	27.0 (22)	15.0 (22)	34.0 (68)
Mixed (*n* = 26)	19.5 (19)	23.0 (34)	34.5 (114)
6. Magnet strength			
1.5T (*n* = 85)	22.0 (18)	18.0 (25)	58.0 (102)
2T (*n* = 2)	10.5	23.0	792.0 (158)
3T (*n* = 325)	25.0 (23)	17.0 (26)	25.0 (56)
4T (*n* = 4)	27.5 (29)	12.0 (35)	63.0 (103)
7T (*n* = 1)	13	18	32
Unknown (*n* = 9)	30.0 (32)	10.0 (21)	58.0 (149)
7. Major fMRI software			
SPM (*n* = 256)	24.0 (20)	17.0 (23)	40.0 (89)
AFNI (*n* = 28)	23.5 (23)	21.0 (30)	23.5 (85)
FSL (*n* = 41)	26.0 (29)	18.0 (32)	16.0 (36)
BrainVoyager (*n* = 35)	20.0 (25)	17.0 (33)	30.0 (60)
Others (*n* = 7)	12.0 (35)	32.0 (47)	130.0 (333)
Mixed (*n* = 51)	30.0 (39)	16.0 (24)	27.0 (48)
Unknown (*n* = 8)	15.5 (27)	4.0 (26)	117.5 (198)
8. Whole brain analysis			
Yes (*n* = 323)	24.0 (21)	18.0 (25)	34.0 (73)
No, ROI analysis only (*n* = 101)	24.0 (20)	16.0 (21)	28.0 (79)
Unknown (*n* = 2)	35.5	30.0	218.5
9. Statistical correction			
FWE (*n* = 243)	26.0 (24)	19.0 (24)	22.0 (55)
FDR (*n* = 73)	26.0 (27)	16.0 (15)	47.0 (85)
Unclear method (*n* = 39)	20.0 (20)	18.0 (26)	24.0 (99)
Uncorrected (*n* = 55)	20.0 (14)	15.0 (27)	55.0 (117)
Unknown (*n* = 16)	15.5 (25)	0 (7)	70.0 (91)
10. Connectivity analysis			
None (*n* = 370)	24.0 (22)	17.0 (24)	35.0 (85)
Yes (*n* = 56)	25.5 (24)	19.0 (25)	21.0 (37)

(Continues)

TABLE 1 (Continued)

1. Obesity-related (21.3% of 426 articles, *n* = 91; obesity = 68, antiobesity drug = 6, obesity treatment = 17).
2. Drinking-related (12.2%, *n* = 52; drinking = 41, antidrinking drug = 6, drinking treatment = 5).
3. Eating disorder-related (8.2%, *n* = 35; eating disorder = 33, emotional eating = 2).
4. Psychiatric-related (3.1%, *n* = 13; depression = 3, depression drug = 2, antipsychotic drug = 1, autism = 2, neuroticism = 2, Parkinson's disease drug = 1, dopamine drug = 1, schizophrenia = 1).

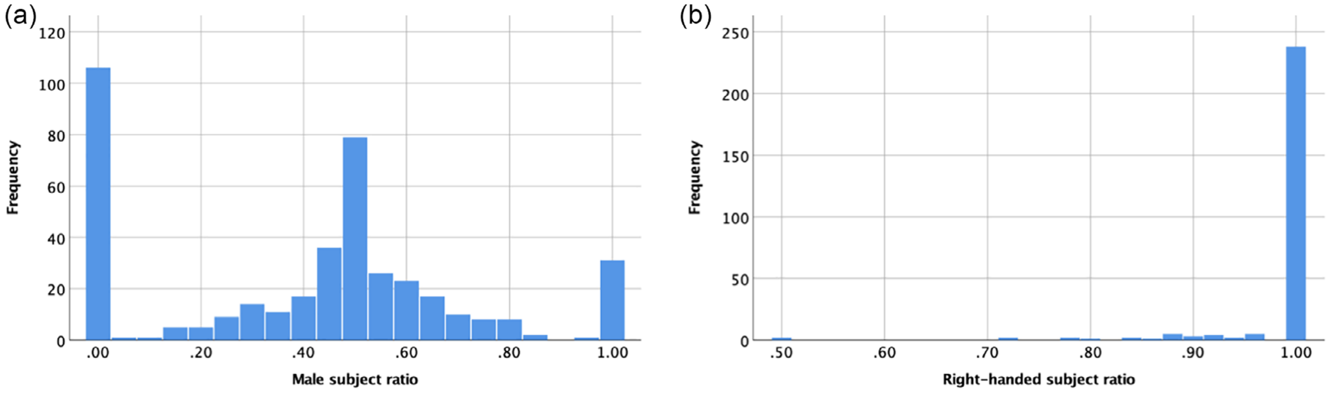


FIGURE 2 The frequency distribution of (a) male subject ratio and (b) right-handed subject ratio. The articles without these pieces of information are not shown in the charts. Binning was applied for simpler illustrations. For example, there is a bar in the middle of panel (a) with frequency = 79, which covers male subject ratio from 0.475 to 0.525 (*n* = 10 for 0.48; *n* = 7 for 0.49; *n* = 53 for 0.50; *n* = 4 for 0.51; and *n* = 5 for 0.52)

5. syndrome-related (2.6%, *n* = 11; Prader-Willi syndrome = 4, polycystic ovary syndrome = 2, metabolic syndrome = 2, diabetes = 1, diabetes treatment = 2).
6. Others (1.4%, *n* = 6; blindness = 1, taste disorder = 1, olfactory impairment = 1, smoking = 2, cocaine abuse = 1).

Over half of the studies (50.5%) used visual stimuli, followed by tastants (30.8%) (Table 1). The studies were mainly dealing with hedonics (43.0%) and somatosensory (37.6%) aspects, used 3T MRI scanner (76.3%) and SPM software (60.1%). Around 24% of the studies reported only ROI analysis without whole brain analysis, and FWE correction was the most popular correction method for multiple comparisons (57.0%). Only 13.1% of studies reported analysis of brain connectivity.

3.2 | Longitudinal results (time trends)

Since the late 1990s, the literature has been growing steadily. The growth accelerated and reached a peak at around 2012–2013 with around 50 papers published each year. Since then, the annual productivity gradually returned to the level of the mid 2000s (Figure 3). The median sample size significantly increased from 11.5 in 1998 to 35.5 in 2019 (rho = 0.387, *p* < .001, Figure 4a). In particular, year 2009 marked the year when studies with over 100 subjects were published, and since then the median sample size has always been above 20. Meanwhile, the papers usually reported around 10–30 foci in their main text, which the median did not show a significant time trend (rho = −0.042, *p* = .386, Figures 4b). On the other hand, the median journal impact factor and the normalized number of citations (citations per year) of the papers showed a significant decreasing trend over the last two decades (rho = −0.118, *p* = .015; and rho = −0.632, *p* < .001, Figure 4c,d). Both metrics peaked at 2002–2003 and then gradually decreased. Meanwhile, the median male subject ratio has been stable at 0.4–0.6 (rho = −0.015, *p* = .760, Figure 4e).

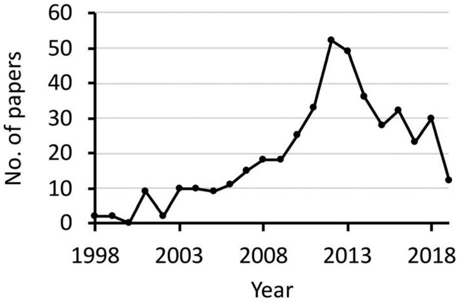
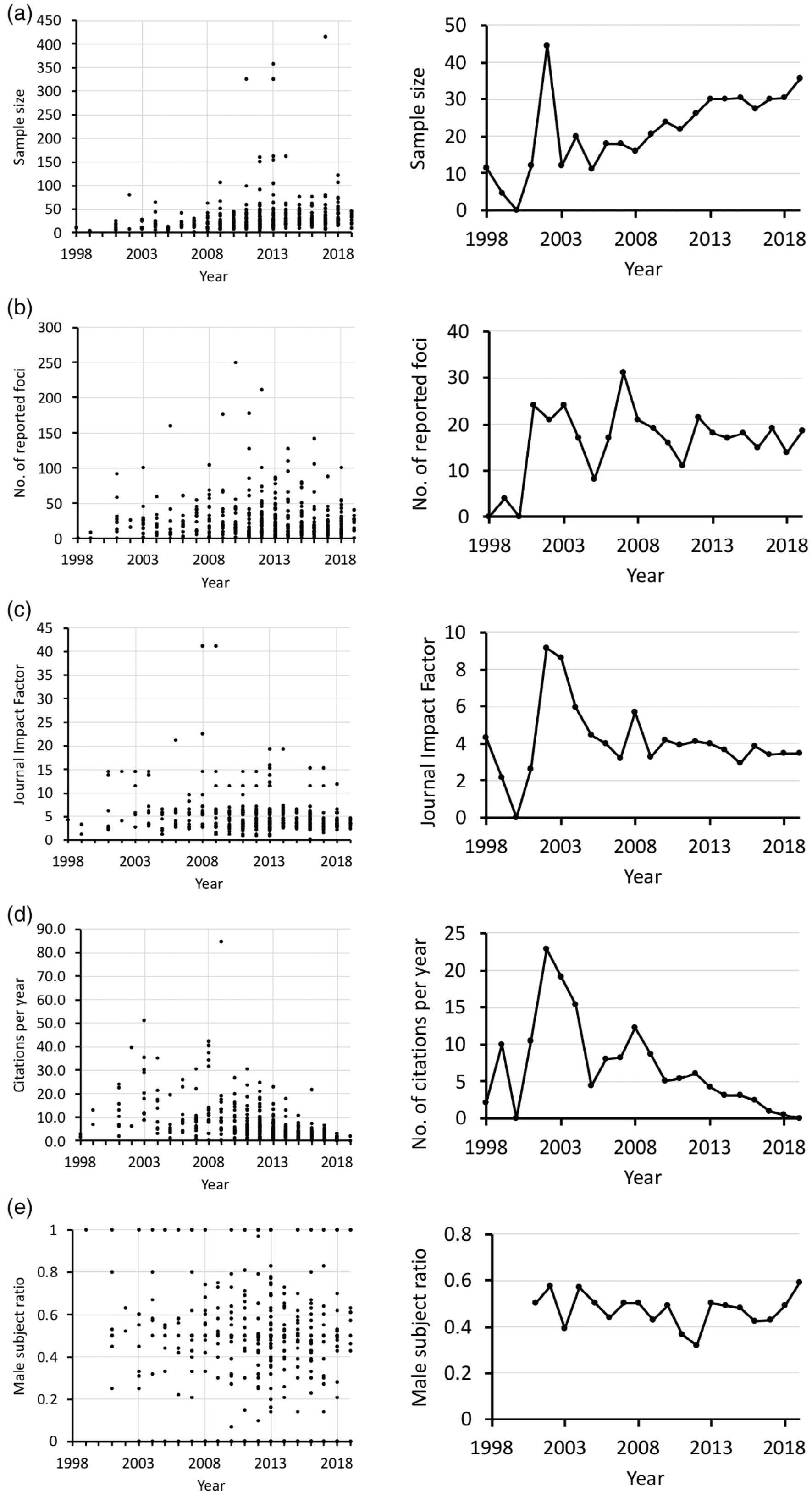


FIGURE 3 Annual publication count of food and taste fMRI papers. fMRI, functional magnetic resonance imaging

When the sex composition of the subjects was viewed categorically, we found that the all females papers were nearly always more prevalent than the all males papers since 2003, except in 2005, 2007, and 2019 (Figure 5a). This was reflected by the fact that all females' papers were thrice as much as all males' papers in overall. For other categorical variables, we noticed that there seemed to be a decreasing trend of papers with mixed handedness (Figure 5b). However, every year around one-fourth to one-half of the papers did not report handedness. For subject type, papers recruiting only patients started to appear in 2007, and now constituted 13.4% of all papers (Figure 5c). For health problems investigated, drinking-related issues had been of huge interest since the early 2000s, but more recently there were increasing interests of eating disorders and obesity (Figure 5d). Other syndrome-related and psychiatric issues constituted 5.4% of the literature only. Meanwhile, visual stimulus was the most popular all-time stimuli type (Figure 5e). Together with tastants, they were accounted for 80% of all papers in overall, and nearly in every year. Odor, auditory, and electric taste constituted a minority. The papers were mainly investigating hedonics and somatosensory aspects of food and taste processing, for which the proportions were quite stable over the years (Figure 5f).

FIGURE 4 Time trends of (a) sample size; (b) number of foci reported in the main text; (c) journal impact factor; (d) citations per year; and (e) male subject ratio of food and taste fMRI papers. The line charts present the trends of the median. fMRI, functional magnetic resonance imaging



For fMRI-specific parameters, we observed a gradual decline in the popularity of 1.5T scanners, to be replaced by 3T scanners (Figure 5g). The use of 2T, 4T, and 7T scanners were scarce. Meanwhile, SPM has been the dominant processing software with over 60% overall share (Figure 5h). In the 2010s, there has been more papers using multiple softwares (mixed), mostly with SPM doing the pre-processing and AFNI determining the statistical threshold. Whole brain analysis has been often reported in the papers (Figure 5i). There

FIGURE 5 Time trends of (a) sex composition; (b) handedness; (c) subject type; (d) health problem investigated; (e) stimuli type; (f) investigation nature; (g) magnet strength; (h) fMRI software; (i) reporting of whole brain analysis; (j) statistical correction method; and (k) reporting of connectivity analysis of food and taste fMRI papers. The bar charts show the annual data, whereas the pie charts show an overall summary. fMRI, functional magnetic resonance imaging

FIGURE 5 (Continued)

seemed to be no increasing trend of papers reporting only ROI analysis since 2001. Regarding methods for statistical correction, FWE has been popular (57.0%) all the time (Figure 5j). FDR (17.1%) began to take a share in the mid-2000s, rose to maximum popularity in 2010, and since then kept on a decreasing track. In overall, 12.9% of papers reported uncorrected statistics only. Finally, connectivity analysis has been reported since 2008, and recently around 20% of papers published each year reported some forms of connectivity analysis

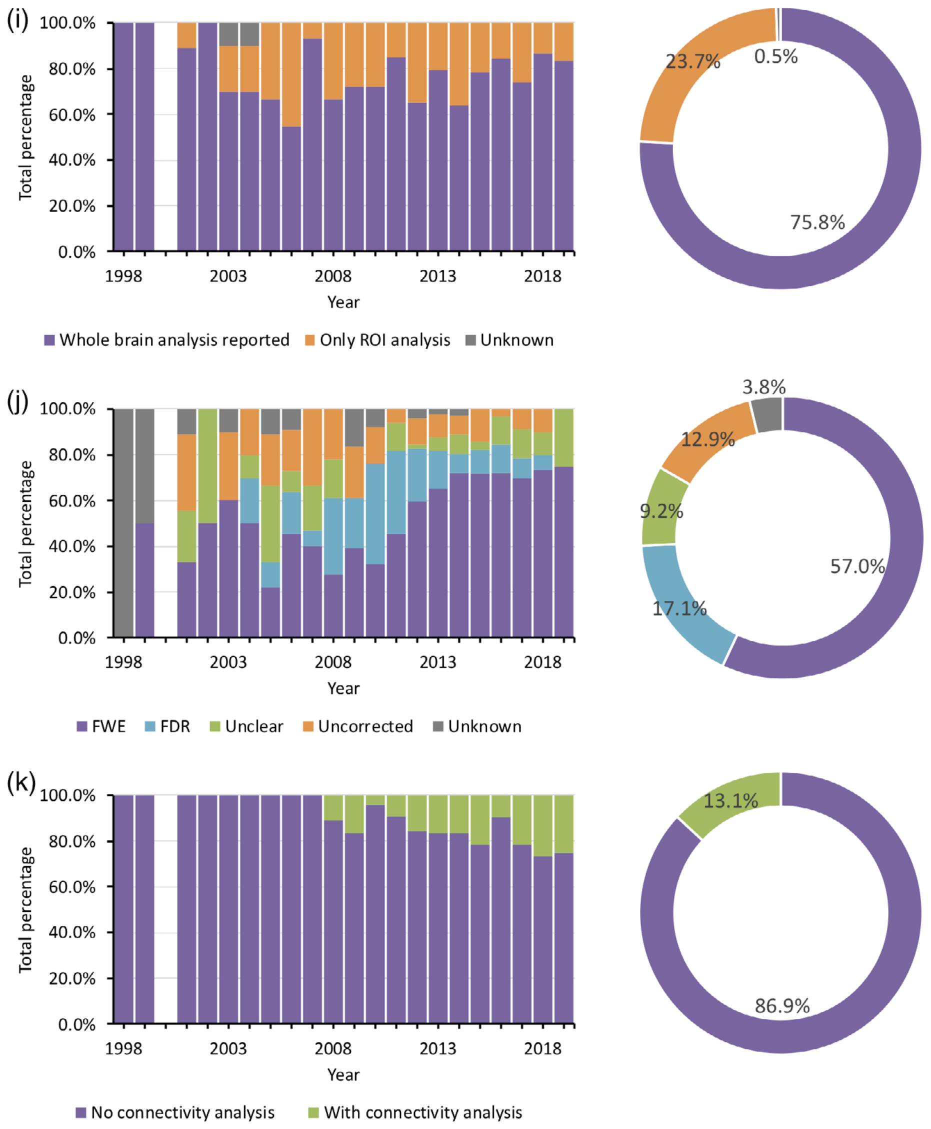


FIGURE 5 (Continued)

(Figure 5k). Because of complexity, we did not conduct statistical tests on the time trends of these categorical variables.

3.3 | Association between citation count and potential influencing factors

Sample size and publication year had a negative correlation to the number of citations and journal impact factor had a positive correlation (Table 2 and Figure 6). Moreover, magnet strength, fMRI software, statistical correction method, and involvement of connectivity analysis were significant influencing factors. In particular, the median citation count of studies using a 1.5T scanner was higher than that using a 3T scanner. Studies reporting statistics with FWE correction had a higher median citation count than those with FDR and uncorrected statistics. Meanwhile, studies without connectivity analysis had a higher median citation count than their counterparts with connectivity analysis.

Male and right-handed subject ratios did not significantly affect the citation count. Additional exploratory analysis was conducted to see if the citation count and journal impact factor differed between uni-sex versus mixed/unknown subjects, between all right-handed versus mixed/unknown subjects. Mann–Whitney test was conducted. Both sex and handedness did not affect the citation count (U = 19,189.5, p = .609; and U = 22,123.0, p = .844). Similarly, they did not affect the journal impact factor (U = 19,422.0, p = .752; and U = 20,974.0, p = .268).

3.4 | Other associations between the extracted parameters

There was no significant influencing factor of the number of reported foci in the main text (Table 3). Meanwhile, sample size positively correlated with journal impact factor (rho = 0.203, p < .001). This

TABLE 2 Potential influencing factors of the number of citations received by the food and taste fMRI papers

Factor	Statistical test	Statistic value	*p* Value
Sample size	Spearman correlation	rho = −0.166	.001*
No. of foci reported in main text	Spearman correlation	rho = 0.071	.142
Publication year	Spearman correlation	rho = −0.814	<.001*
Male subject ratio	Spearman correlation	rho = −0.046	.348
Right-handed subject ratio	Spearman correlation	rho = −0.035	.569
2018 Journal Impact Factor	Spearman correlation	rho = 0.372	<.001*
Subject type	Kruskal-Wallis test	*H* = 5.344	.069
Health problem focused	Kruskal-Wallis test	*H* = 6.392	.381
Stimuli type	Kruskal-Wallis test	*H* = 10.274	.068
Investigation nature	Kruskal-Wallis test	*H* = 1.320	.724
Magnet strength	Kruskal-Wallis test	*H* = 26.006	<.001*,[a]
fMRI software	Kruskal-Wallis test	*H* = 22.849	.001*,[b]
Whole brain versus ROI analysis	Mann–Whitney test	*U* = 16,092.5	.839
Statistical correction	Kruskal-Wallis test	*H* = 22.508	<.001*,[c]
Connectivity analysis involvement	Mann–Whitney test	*U* = 7,997.0	.006*,[d]

Notes: See Table 1 for detailed citation counts. For magnet strength, fMRI software, whole brain versus ROI analysis and statistical correction, papers with unknown information were excluded from the tests.

[a] 1.5T > 3T.

[b] Others > FSL.

[c] FWE > FDR = uncorrected.

[d] No connectivity analysis > With connectivity analysis. * means p equal or smaller than (symbol) 0.001.

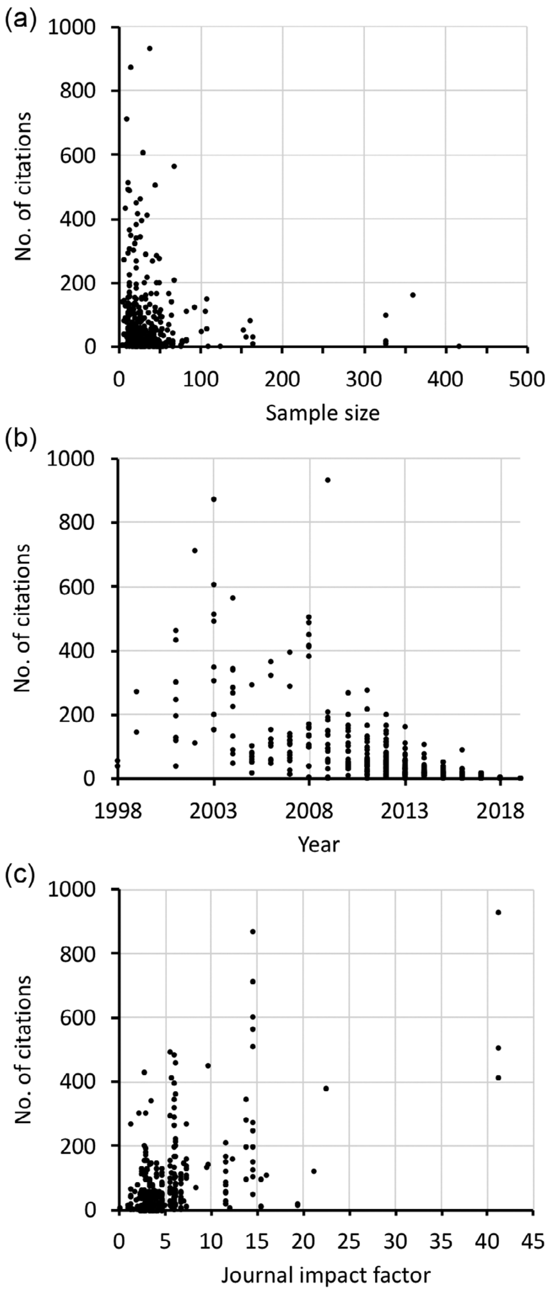


FIGURE 6 Scatter plots of the number of citations against (a) sample size, (b) publication year, and (c) journal impact factor

correlation became insignificant after correcting for publication year (rho = 0.661, *p* = .540). The median sample size of mixed handedness studies (25.0) was not statistically larger than the all right-handed (23.5) and unknown (28.0) counterparts (*H* = 5.201, *p* = .074).

4 | DISCUSSION

This study surveyed over 400 fMRI articles on food and taste research so as to report the time trends of experimental designs including fMRI-specific parameters. Results showed that several parameters, for example, publication count, sample size, normalized citations and journal impact factor, have evolved over the last two decades. Moreover, we have identified several influencing factors of the citation count, but not the number of foci.

4.1 | Potential reporting bias

This study has revealed no statistically significant relationship between sample size and the number of foci for food and taste fMRI articles. This finding is in line with previous reports on neuroimaging studies of sex differences (179 original studies) (David et al., 2018), and fMRI studies in general (1,778 original studies) (David et al.,

TABLE 3 Potential influencing factors of the number of foci reported in the main text of the food and taste fMRI papers

Factor	Statistical test	Statistic value	*p* Value
Sample size	Spearman correlation	rho = −0.005	.911
Publication year	Spearman correlation	rho = −0.042	.386
Male subject ratio	Spearman correlation	rho = −0.079	.111
Right-handed subject ratio	Spearman correlation	rho = 0.006	.920
No. of Web of Science citations	Spearman correlation	rho = 0.071	.142
2018 Journal Impact Factor	Spearman correlation	rho = −0.091	.060
Subject type	Kruskal-Wallis test	*H* = 0.055	.973
Healthy problem focused	Kruskal-Wallis test	*H* = 8.982	.175
Stimuli type	Kruskal-Wallis test	*H* = 3.585	.611
Investigation nature	Kruskal-Wallis test	*H* = 1.278	.734
Magnet strength	Kruskal-Wallis test	*H* = 1.922	.750
fMRI software	Kruskal-Wallis test	*H* = 4.591	.468
Whole brain versus ROI analysis	Mann–Whitney test	*U* = 14,214.5	.051
Statistical correction	Kruskal-Wallis test	*H* = 1.804	.614
Connectivity analysis involvement	Mann–Whitney test	*U* = 9,390.5	.259

Notes: See Table 1 for detailed counts on the number of foci. For magnet strength, fMRI software, whole brain versus ROI analysis and statistical correction, papers with unknown information were excluded from the tests.
Abbreviation: fMRI, functional magnetic resonance imaging.

2013); though a report on voxel-based morphometry studies of psychiatric and neurological disorders found that the number of foci increased 2% for every 10-patient increase in the sample size (324 original studies) (Fusar-Poli et al., 2014). Meanwhile, we found that studies with unknown type of multiple comparison correction method tended to report fewer foci relative to other studies. But that was due to the fact that most of those studies reported ROI analysis only, and they extracted the parameter estimates or signal change across the whole pre-defined anatomical structures without applying for a threshold to exclude individual voxels. Otherwise, we did not find a moderation by the type of multiple comparison correction to the number of reported foci. These findings may continue to support the notion that there seems to be a potential reporting bias and significance bias, considered together with the finding that studies with larger samples did not tend to be cited more than those with smaller samples.

4.2 | Dominance of right-handed subjects and sample representativeness

A recent survey, based on approximately 1,000 articles, reported that only 3–4% of over 30,000 research subjects recruited for neuroimaging experiments were left-handed or ambidextrous, compared to their prevalence of 10–13% in the general population (Bailey et al., 2019). In this dataset, we also found that the subjects were mainly right-handed. For articles that reported handedness of subjects, 89.1% of them recruited only right-handed subjects. However, it should also be noted that 37.3% of the analyzed articles did not report subject handedness – this ratio is considerably much larger than the 9.2% of sensorimotor-related articles as reported by Bailey et al. (2019).

In general, there is always the issue of homogeneity versus representativeness to be considered at the stage of subject recruitment. There is no simple answer to this issue. For a general research study, a highly homogeneous sample (e.g., with limited age range) would lead to limitations that inferences may not be drawn for subjects outside the inclusion range (Rothman, Gallacher, & Hatch, 2013). To be representative, it usually requires a much larger sample, be it a sample recruited according to the proportions in the population, or groups of samples in similar sizes (Richiardi, Pizzi, & Pearce, 2013; Rothman et al., 2013). Whereas an efficacy study may recruit a more homogeneous sample, a representative sample is more suited for an effectiveness study (Glasgow, Lichtenstein, & Marcus, 2003).

The relevance of handedness to gustatory neuroimaging studies was reported by pioneer works by Cerf, Lebihan, Van de Moortele, Mac Leod, and Faurion (1998) and Faurion et al. (1999). These two fMRI studies recruited 10 healthy young adults, 5 right-handed and 5 left-handed, and found that the inferior insula predominantly activated, in response to gustatory stimulations, on the contralateral side relative to the dominant hand (Cerf et al., 1998; Faurion et al., 1999). This issue was reiterated one decade later, when Wakita et al. (2009) reported that the right-handed subjects had a significant morphometric asymmetry in the primary gustatory area in the insula compared to the left-handed subjects, and reasoned that the presence of motor speech area in the language-dominant hemisphere shifted the primary gustatory area posteriorly. Perhaps these seminal works have had a larger influence on the succeeding literature. In addition, a recent food-related neuroimaging guideline advocated for the reporting of handedness, and accounting for non-right-handedness in the analyses, implying that subjects should be right-handed by default (Smeets et al., 2019). Consistent to our past experience in fMRI research, perhaps it is a consensus in the gustatory neuroimaging community to recruit right-handed subjects as a norm, so that only two of the surveyed articles explained why they recruited all right-handed subjects. The question of whether the majority of these gustatory fMRI studies, based on all or nearly all right-handed subjects, can be generalized to the left-handed population, remains to be debated.

4.3 | Relevance of subject sex

Two-thirds of the surveyed articles that recruited a uni-sex sample did not explain why they did not recruit subjects from both sexes. In neuroscience research using animal models, nearly half of the studies were male only, with 25% used both sexes and 9% female only (Beery, 2018). By surveying neuroscience literature (both human and animal studies) published between 2010 and 2014, Will et al. (2017) found that increasing numbers of articles reported sex, but sex bias remains present, with increasing numbers of articles reported the sole use of males. However, a meta-analysis pointed out that female rats were not more variable at any time of the estrous cycle compared to male rats, and that the neuroscientific findings should still be valid even if the estrous cycle was not accounted for (Becker, Prendergast, & Liang, 2016). Indeed, National Institutes of Health (NIH) grants request applicants to account for sex as a biological variable, meaning that subjects from both sexes should be recruited, unless a strong justification is made (Shansky & Woolley, 2016). In particular, the presence (or absence) of sex differences and sex specific effects should be investigated and reported (Shansky & Woolley, 2016). Interestingly, our survey found that the food and taste literature was biased toward the females instead of the males. This could be attributed to the investigation of eating disorders, such as anorexia and bulimia, which affect predominantly the females. This female predominance is similar to the research literature in behavioral physiology, endocrinology, and reproduction (Beery & Zucker, 2011).

A recent systematic review by Chao et al. (2017) reported that women in hungry state showed heightened activation in the frontal, limbic, striatal areas, and fusiform gyrus compared to men. Meanwhile, a meta-analysis reported that men had larger responses in the cingulate cortex upon receiving food or eating stimuli, whereas women had larger responses in the parahippocampus, the thalamus, and the precuneus (Yeung, 2018a). Depending on the taste of the tastants, Haase, Green, and Murphy (2011) reported 2–11 foci showing differential activation between males and females. All these pieces of evidence imply that there exist sexual differences in the brain responses to food and taste stimuli, which should be accounted for by recruiting subjects from both sexes so that the results can be better generalized to the population. Meanwhile, our results showed that the most common reason for recruiting an all-male sample was to avoid the variability/effects of menstrual cycle, whereas that for recruiting an all-female sample was that eating disorder is predominant in females. The latter reason tended to apply to studies that recruited patients rather than healthy subjects. Besides, it seemed that the majority of the studies recruiting both males and females in our analyzed dataset actually recruited both sexes in a relatively balanced ratio, as illustrated by Figure 2a.

4.4 | Clear description of statistical correction is recommended

Approximately 10% of the surveyed articles did not provide a clear description on how they corrected for multiple comparisons. This phenomenon mainly occurred in articles using FSL software for data processing. A typical example of the description was: "maps were corrected for multiple comparisons using cluster-based thresholding ($z > 1.96$, cluster $p < .05$)." FSL reports FWE-corrected p values for cluster-level inference by default, but other correction methods are also available (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012). Therefore, a clear and explicit statement on how the statistics are corrected is recommended.

It seemed that FDR correction has experienced a rise in popularity in the 2000s until 2010 and then declined gradually. In 2009 and 2010, Chumbly, Friston and colleagues published two renowned papers with a statement that inferences based on cluster-level FDR correction are acceptable but voxel-wise FDR procedures are unacceptable by causing a substantial increase in false positives (J. Chumbley, Worsley, Flandin, & Friston, 2010; J. R. Chumbley & Friston, 2009). Although the current study did not distinguish voxel-wise from cluster-wise inference, the decline of FDR correction seemed to be consistent, right after the publication of these papers. On the other hand, we found that 12.9% of the studies only reported statistics without correction for multiple comparisons. This ratio (12.9%) is larger than the 4.4% reported by a survey on fMRI papers solely published in the year 2017 (Yeung, 2018b), but comparable to the figures reported by the preceding reports (6.0–40.9%) (Carp, 2012; Guo et al., 2014; Woo, Krishnan, & Wager, 2014).

4.5 | Increasing sample size over time

The median sample size of the analyzed dataset increased from around 10 in the late 1990s and early 2000s, to around 20 in the late 2000s, and further to over 30 in the late 2010s, reaching 35.5 in 2019. This upward trend is comparable to Poldrack et al. (2017), who reported a similar trend from around 10 subjects in the early 2000s, up to 28.5 as of 2015. It is also consistent to Yeung (2018b), who reported a median of 33 for papers published in 2017.

4.6 | Other comments and future perspective

Unlike the visual and the sensorimotor systems, where the retinotopy and cortical homunculus are extensively studied and well-known, there have been few studies in the gustatory neuroscience field that attempted to map the differential cortical representation of the basic tastes in humans. In 2009, Schoenfeld et al. (2004) recruited six subjects for an fMRI study with a 1.5T scanner, and revealed that the five basic tastes tended to be chemotopically represented in the insula, with sweet and salty tastes represented more anteriorly relative to umami, sour, and bitter tastes. Seven years later, Chen, Gabitto, Peng, Ryba, and Zuker (2011) succeeded in using two-photon calcium imaging to identify discrete clusters of neurons responsive to sweet, bitter, umami, and salty tastes in the rat gustatory cortex. Precisely, sweet taste was represented more rostrally to salty and umami tastes, which were in turn rostral to bitter taste. The second human fMRI study took place in 2017, with ten subjects scanned with a 1.5T scanner, and found that saltiness, sourness, and umami were represented closely in a cluster in the anterior insula, as compared to bitterness and sweetness in another cluster in the posterior insula (Prinster et al., 2017). Very recently, Chikazoe, Lee, Kriegeskorte, and Anderson (2019) scanned 20 adults with a 3T scanner and 11 adults with a 7T scanner, and their multivoxel pattern analysis found that the activation pattern in the anterior to middle insula was discriminative of sour, sweet, bitter, and salty tastes. However, they did not find discrete, non-overlapping clusters of activation representing each taste. It remains to be elucidated if a clear gustotopic map exists in human insula.

Meanwhile, the majority of the food and taste fMRI studies was focused on brain mapping, including correlational analyses between brain activity level and behavioral/physiological scores (e.g., ratings of perceived intensity, pleasantness, wanting; BMI; scores from psychosocial scales). We anticipate that, after understanding the neurobiological basis of cerebral processing of food and taste stimuli, future studies will focus more on the functional and effective connectivity, as connectivity studies have been on the rise (Friston, 2011; Poldrack, 2012). Future fMRI studies may use scanner with stronger magnets, such as 7T, to increase the signal-to-noise ratio. However, it was reported that many subjects felt a metallic taste in their mouth upon mild head movements during scanning inside a 7T MRI (Cavin, Glover, Bowtell, & Gowland, 2007). This issue may potentially confound future fMRI studies using 7T MRI.

5 | CONCLUSION

This survey, using food and taste fMRI studies as a primer, confirmed the previous finding that there existed no correlation between sample size and number of foci reported, implying a potential reporting bias in the fMRI literature. It is reassuring to see that the median sample size has been increasing gradually over the last two decades, implying a potential improvement in the power of the studies. The dominance of right-handed subjects without explicit elaboration, and the prevalence of studies recruiting a uni-sex sample should be made aware to the research community and better accounted for. Besides, details about the methods of statistical correction should be disclosed more to allow a better understanding and replication of the results. Finally, there should be an increasing focus on connectivity analysis.

ACKNOWLEDGMENTS

SBE is supported by the Deutsche Forschungsgemeinschaft (DFG, EI 816/11-1), the National Institute of Mental Health (R01-MH074457), the Helmholtz Portfolio Theme “Supercomputing and Modeling for the Human Brain,” and the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 785907 (HBP SGA2).

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT

The raw data of the articles can be found in Supplementary Data.

ORCID

Andy W. K. Yeung https://orcid.org/0000-0003-3672-357X

REFERENCES

Bailey, L. M., McMillan, L. E., & Newman, A. J. (2019). A sinister subject: Quantifying handedness-based recruitment biases in current neuroimaging research. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.14542. [Epub ahead of print]

Becker, J. B., Prendergast, B. J., & Liang, J. W. (2016). Female rats are not more variable than male rats: a meta-analysis of neuroscience studies. *Biology of Sex Differences*, *7*(1), 34.

Beery, A. K. (2018). Inclusion of females does not increase variability in rodent research studies. *Current Opinion in Behavioral Sciences*, *23*, 143–149.

Beery, A. K., & Zucker, I. (2011). Sex bias in neuroscience and biomedical research. *Neuroscience & Biobehavioral Reviews*, *35*(3), 565–572.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

Carp, J. (2012). The secret lives of experiments: methods reporting in the fMRI literature. *NeuroImage*, *63*(1), 289–300.

Cavin, I. D., Glover, P. M., Bowtell, R. W., & Gowland, P. A. (2007). Thresholds for perceiving metallic taste at high magnetic field. *Journal of Magnetic Resonance Imaging*, *26*(5), 1357–1361.

Cerf, B., Lebihan, D., Van de Moortele, P., Mac Leod, P., & Faurion, A. (1998). Functional lateralization of human gustatory cortex related to handedness disclosed by fMRI study. *Annals of the New York Academy of Sciences*, *855*(1), 575–578.

Chao, A. M., Loughead, J., Bakizada, Z. M., Hopkins, C. M., Geliebter, A., Gur, R. C., & Wadden, T. A. (2017). Sex/gender differences in neural correlates of food stimuli: a systematic review of functional neuroimaging studies. *Obesity Reviews*, *18*(6), 687–699.

Chen, X., Gabitto, M., Peng, Y., Ryba, N. J., & Zuker, C. S. (2011). A gustotopic map of taste qualities in the mammalian brain. *Science*, *333*(6047), 1262–1266.

Chikazoe, J., Lee, D. H., Kriegeskorte, N., & Anderson, A. K. (2019). Distinct representations of basic taste qualities in human gustatory cortex. *Nature Communications*, *10*(1), 1048.

Chumbley, J., Worsley, K., Flandin, G., & Friston, K. (2010). Topological FDR for neuroimaging. *NeuroImage*, *49*(4), 3057–3064.

Chumbley, J. R., & Friston, K. J. (2009). False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage*, *44*(1), 62–70.

David, S. P., Naudet, F., Laude, J., Radua, J., Fusar-Poli, P., Chu, I., … Ioannidis, J. P. (2018). Potential reporting bias in neuroimaging studies of sex differences. *Scientific Reports*, *8*(1), 6082.

David, S. P., Ware, J. J., Chu, I. M., Loftus, P. D., Fusar-Poli, P., Radua, J., … Ioannidis, J. P. (2013). Potential reporting bias in fMRI studies of the brain. *PLoS One*, *8*(7), e70104.

Faurion, A., Cerf, B., Van De Moortele, P.-F., Lobel, E., Mac Leod, P., & Le Bihan, D. (1999). Human taste cortical areas studied with functional magnetic resonance imaging: evidence of functional lateralization related to handedness. *Neuroscience Letters*, *277*(3), 189–192.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connectivity*, *1*(1), 13–36.

Fusar-Poli, P., Radua, J., Frascarelli, M., Mechelli, A., Borgwardt, S., Di Fabio, F., … David, S. P. (2014). Evidence of reporting biases in voxel-based morphometry (VBM) studies of psychiatric and neurological disorders. *Human Brain Mapping*, *35*(7), 3052–3065.

Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health*, *93*(8), 1261–1267.

Guo, Q., Parlar, M., Truong, W., Hall, G., Thabane, L., McKinnon, M., … Pullenayegum, E. (2014). The reporting of observational clinical functional magnetic resonance imaging studies: a systematic review. *PLoS One*, *9*(4), e94412.

Haase, L., Green, E., & Murphy, C. (2011). Males and females show differential brain activation to taste when hungry and sated in gustatory and reward areas. *Appetite*, *57*(2), 421–434.

Huerta, C. I., Sarkar, P. R., Duong, T. Q., Laird, A. R., & Fox, P. T. (2014). Neural bases of food perception: Coordinate-based meta-analyses of neuroimaging studies in multiple modalities. *Obesity*, *22*(6), 1439–1446.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *Neuroimage*, *62*(2), 782–790.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*, *7*(5), 604–609.

Poldrack, R. A. (2012). The future of fMRI in cognitive neuroscience. *NeuroImage*, *62*(2), 1216–1220.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., … Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115–126.

Prinster, A., Cantone, E., Verlezza, V., Magliulo, M., Sarnelli, G., Iengo, M., … Esposito, F. (2017). Cortical representation of different taste modalities on the gustatory cortex: A pilot study. *PLoS One*, *12*(12), e0190164.

Richiardi, L., Pizzi, C., & Pearce, N. (2013). Commentary: Representativeness is usually not necessary and often should be avoided. *International Journal of Epidemiology*, *42*(4), 1018–1022.

Rothman, K. J., Gallacher, J. E., & Hatch, E. E. (2013). Why representativeness should be avoided. *International Journal of Epidemiology*, *42*(4), 1012–1014.

Schoenfeld, M., Neuer, G., Tempelmann, C., Schüssler, K., Noesselt, T., Hopf, J.-M., & Heinze, H.-J. (2004). Functional magnetic resonance tomography correlates of taste perception in the human primary taste cortex. *Neuroscience*, *127*(2), 347–353.

Shansky, R. M., & Woolley, C. S. (2016). Considering sex as a biological variable will be valuable for neuroscience research. *Journal of Neuroscience*, *36*(47), 11817–11822.

Smeets, P. A., Dagher, A., Hare, T. A., Kullmann, S., van der Laan, L. N., Poldrack, R. A., … Veldhuizen, M. G. (2019). Good practice in food-related neuroimaging. *American Journal of Clinical Nutrition*, *109*(3), 491–503.

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications Biology*, *1*(1), 62.

Veldhuizen, M. G., Albrecht, J., Zelano, C., Boesveldt, S., Breslin, P., & Lundström, J. N. (2011). Identification of human gustatory cortex by activation likelihood estimation. *Human Brain Mapping*, *32*(12), 2256–2266.

Wakita, M., Kobayakawa, T., Saito, S., Sakai, N., Hiai, Y., Hirai, T., … Ogawa, H. (2009). Handedness: dependent asymmetrical location of the human primary gustatory area, area G. *Neuroreport*, *20*(4), 450–455.

Will, T. R., Proaño, S. B., Thomas, A. M., Kunz, L. M., Thompson, K. C., Ginnari, L. A., … Dorris, D. M. (2017). Problems and progress regarding sex bias and omission in neuroscience research. *eNeuro*, *4*(6), ENEURO.0278-0217.2017.

Woo, C.-W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. *NeuroImage*, *91*, 412–419.

Yeung, A. W. K. (2018a). Sex differences in brain responses to food stimuli: a meta-analysis on neuroimaging studies. *Obesity Reviews*, *19*(8), 1110–1115.

Yeung, A. W. K. (2018b). An updated survey on statistical thresholding and sample size of fMRI studies. *Frontiers in Human Neuroscience*, *12*, 16.

Yeung, A. W. K., Goto, T. K., & Leung, W. K. (2017). Basic taste processing recruits bilateral anteroventral and middle dorsal insulae: an activation likelihood estimation meta-analysis of fMRI studies. *Brain and Behavior*, *7*, e00655.

Yeung, A. W. K., Goto, T. K., & Leung, W. K. (2018). Affective value, intensity and quality of liquid tastants/food discernment in the human brain: An activation likelihood estimation meta-analysis. *NeuroImage*, *169*, 189–199. https://doi.org/10.1016/j.neuroimage.2017.12.034

Yeung, A. W. K., Wong, N. S. M., Lau, H., & Eickhoff, S. B. (2019). Human brain responses to gustatory and food stimuli: A meta-evaluation of neuroimaging meta-analyses. *NeuroImage*, *202*, 116111.

SUPPORTING INFORMATION

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

How to cite this article: Yeung AWK, Wong NSM, Eickhoff SB. Empirical assessment of changing sample-characteristics in task-fMRI over two decades: An example from gustatory and food studies. *Hum Brain Mapp*. 2020;41:2460–2473. https://doi.org/10.1002/hbm.24957