

## REVIEW ARTICLE OPEN



# Automatic, wearable-based, in-field eating detection approaches for public health research: a scoping review

Brooke M. Bell<sup>1</sup>✉, Ridwan Alam<sup>2</sup>, Nabil Alshurafa<sup>3,4</sup>, Edison Thomaz<sup>5</sup>, Abu S. Mondol<sup>6</sup>, Kayla de la Haya<sup>1</sup>, John A. Stankovic<sup>6</sup>, John Lach<sup>7</sup> and Donna Spruijt-Metz<sup>1,8,9</sup>

Dietary intake, eating behaviors, and context are important in chronic disease development, yet our ability to accurately assess these in research settings can be limited by biased traditional self-reporting tools. Objective measurement tools, specifically, wearable sensors, present the opportunity to minimize the major limitations of self-reported eating measures by generating supplementary sensor data that can improve the validity of self-report data in naturalistic settings. This scoping review summarizes the current use of wearable devices/sensors that automatically detect eating-related activity in naturalistic research settings. Five databases were searched in December 2019, and 618 records were retrieved from the literature search. This scoping review included  $N = 40$  studies (from 33 articles) that reported on one or more wearable sensors used to automatically detect eating activity in the field. The majority of studies ( $N = 26$ , 65%) used multi-sensor systems (incorporating  $> 1$  wearable sensors), and accelerometers were the most commonly utilized sensor ( $N = 25$ , 62.5%). All studies ( $N = 40$ , 100.0%) used either self-report or objective ground-truth methods to validate the inferred eating activity detected by the sensor(s). The most frequently reported evaluation metrics were Accuracy ( $N = 12$ ) and F1-score ( $N = 10$ ). This scoping review highlights the current state of wearable sensors' ability to improve upon traditional eating assessment methods by passively detecting eating activity in naturalistic settings, over long periods of time, and with minimal user interaction. A key challenge in this field, wide variation in eating outcome measures and evaluation metrics, demonstrates the need for the development of a standardized form of comparability among sensors/multi-sensor systems and multidisciplinary collaboration.

npj Digital Medicine (2020)3:38; <https://doi.org/10.1038/s41746-020-0246-2>

## INTRODUCTION

Dietary intake (i.e., what and how much is consumed), eating behaviors (i.e., food choices and motives, feeding practices), and context (i.e., who is eating, when, where, with whom, etc.) play a significant role in the development of chronic diseases, including type 2 diabetes, heart disease, and obesity<sup>1–6</sup>. Recent data from the National Health and Nutrition Examination Survey (NHANES) indicate that national obesity prevalence for U.S. adults (39.6%) and for U.S. youth (18.5%) is the highest ever documented (compared with 14.5% and 5.0%, respectively, in the early 1970s)<sup>7–9</sup>. Poor diet is estimated to have contributed to 11 million deaths globally in 2017<sup>10</sup>. Despite this strong association between food intake and health, our ability to accurately assess dietary intake, eating behaviors, and context are three existing challenges in dietary research. Dietary intake/eating behavior assessment in both youth and adult populations historically relies on self-reporting tools<sup>11,12</sup>. The most commonly used tools to assess dietary intake/eating behaviors are 24-h recalls, food records (food diaries), and food frequency questionnaires (FFQ)<sup>13,14</sup>. Major limitations of these methods include participant burden and recall or memory bias<sup>15,16</sup>, which can lead to under- and over-reporting of dietary intake, skewing research findings in both directions<sup>17</sup>. Measurement tools that can minimize these limitations are crucial for

accurately detecting temporal patterns of food and nutrient intake and having measurement sensitivity to detect intake changes, and to ultimately discern the influences of dietary intake and eating behaviors on health outcomes.

Traditional methods of eating assessment typically summarize dietary measures at the hour-, day-, week-, or even year-level<sup>11,14</sup>. Although these can be helpful in understanding relationships between eating behavior and its predictors, important micro-level temporal patterns and processes are not measured nor can they be explored with these measures. The ability to explore micro-level eating activities, such as meal microstructure (the dynamic process of eating, including meal duration, changes in eating rate, chewing frequency, etc.)<sup>18</sup>, food choices<sup>19</sup>, and processes (e.g., eating rate<sup>20</sup>; eating mimicry<sup>21,22</sup>; etc.), is important because recent literature suggests that they may play an important role on food selection, dietary intake, and ultimately, obesity and disease risk.

Emerging technologies present the opportunity to improve these assessment methods by using methods that improve the quality and validity of data that is collected, and by passively measuring eating activity in naturalistic settings over long periods of time with minimal user interaction. Technological advances in dietary assessment tools include: (i) web-based self-administered

<sup>1</sup>Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA. <sup>2</sup>Department of Electrical and Computer Engineering, School of Engineering and Applied Science, University of Virginia, Charlottesville, VA 22904, USA. <sup>3</sup>Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. <sup>4</sup>Department of Computer Science, McCormick School of Engineering, Northwestern University, Chicago, IL 60611, USA. <sup>5</sup>Department of Electrical and Computer Engineering, Cockrell School of Engineering, The University of Texas at Austin, Austin, TX 78712, USA. <sup>6</sup>Department of Computer Science, School of Engineering and Applied Science, University of Virginia, Charlottesville, VA 22904, USA. <sup>7</sup>Department of Electrical and Computer Engineering, School of Engineering and Applied Science, The George Washington University, Washington, DC 20052, USA. <sup>8</sup>Center for Economic and Social Research, Dornsife College of Letters, Arts, and Sciences, University of Southern California, Los Angeles, CA 90089, USA. <sup>9</sup>Department of Psychology, Dornsife College of Letters, Arts, and Sciences, University of Southern California, Los Angeles, CA 90089, USA. ✉email: brooke.bell@usc.edu

24-h recall tool, which aims to reduce respondent burden<sup>23</sup>; (ii) mobile device-assisted ecological momentary assessment (mEMA), which focuses on reducing recall bias by collecting real-time data in naturalistic settings<sup>24</sup>; (iii) photo-assisted and image-based dietary assessments, which attempt to reduce respondent burden and recall bias<sup>25</sup>; and (iv) wearable sensors, which offer a suite of measurement tools that seek to tackle all of these limitations<sup>26,27</sup>. Wearable devices with embedded sensors in particular allow for the passive collection of various data streams that can be used to develop algorithms to infer eating behaviors in naturalistic settings and, with some types of sensors, over long periods of time. Collecting near-continuous data in the context of daily life, where behavior actually occurs, has proven to be extremely difficult for researchers and burdensome for participants<sup>26,27</sup>. However, wearable sensors can lessen the burden by passively collecting data while users go about their daily lives with minimal user input, compared with existing methods.

In the past decade, a range of wearable sensors for the purpose of automating eating detection have been proposed and studied. However, these studies have been primarily conducted in a combination of controlled lab and semi-controlled field settings<sup>28,29</sup>, and for good reason: these systems are challenging to develop, deploy, and evaluate. More recently, however, the research field has rapidly expanded the testing of these devices in the field. Previous research has shown significant differences in eating metrics (e.g., duration of meals, number of bites, etc.) between similar in-lab and in-field studies<sup>30</sup>, illustrating the importance of in-field testing and in-field validation of wearable sensors. Field deployment is crucial because it is where humans are more likely to behave naturally as compared with a research lab setting, and many of the influences on eating behavior cannot be replicated in a laboratory. Moreover, non-eating behavior that confounds these sensors (e.g., smoking, biting nails, etc.) are too many, and not all known, to replicate in a natural way in controlled settings. The data from wearable sensors deployed in the field will offer researchers a wealth of temporally-rich, highly contextualized eating activity data that can address exciting and novel research questions that were previously inexplorable.

The next step toward integrating these wearable technologies into public health research is to continue testing the wearable technology in the field, while beginning to address and solve the unique technical, analytical, and multidisciplinary challenges that arise from this effort. Wearable sensors provide the opportunity to examine and understand real-time eating behavior in context, but their introduction into field testing has been slow and challenging. It is unclear what kind of information is available in the literature about the unique features and challenges of field testing. Moreover, the measures with which these sensors are reported and evaluated is also highly varied and non-uniform. Therefore, a scoping review was conducted in order to:

1. catalog the current use of wearable devices and sensors that automatically detect eating activity (dietary intake and/or eating behavior) specifically in free-living research settings;
2. and identify the sample size, sensor types, ground-truth measures, eating outcomes, and evaluation metrics used to evaluate these sensors.

It was outside the scope of this paper to systematically review contextual factors, if any, detected with these automatic, wearable-based methods; however, a brief discussion on the potential for these methods in future research is included. In addition, a discussion of key challenges of in-field testing and recommendations for future research directions are included.

A scoping review was chosen due to the heterogenous methods used in this field of research.

## METHODS

The reporting of this scoping review adheres to the Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) checklist<sup>31</sup>.

### Information sources

We conducted a literature search of the PubMed, SCOPUS, IEEE Xplore, ACM Digital Library, and Google Scholar (first 50 results) databases for all literature published through December 2019. A research librarian assisted with the literature search strategy, which was further refined by the authorship team. The records retrieved from the databases were imported into a web-based systematic review management tool, Covidence<sup>32</sup>. We also hand searched the reference lists of the originally included publications for additional eligible studies.

### Literature search

The literature search strategy included a combination of keywords and variations to identify articles that addressed (i) eating assessment/detection, (ii) wearable technology, and (iii) free-living research settings. Keywords included “eat”, “food intake”, “diet”; “monitor”, “assess”, “detect”; and “wearable”, “device”, “sensor”, “technology”, “smart watch”, “smartwatch”, “ambulatory”, “free living”, “in field”, “in the wild”. The full search term strategy that was used for each database is outlined in Supplementary Table 1.

### Eligibility criteria

Peer-reviewed journal or conference papers were considered for inclusion in the review if they were published prior to December 22, 2019 and were written in English. Furthermore, eligible papers needed to have described/reported on all of the following components within the contents of the paper:

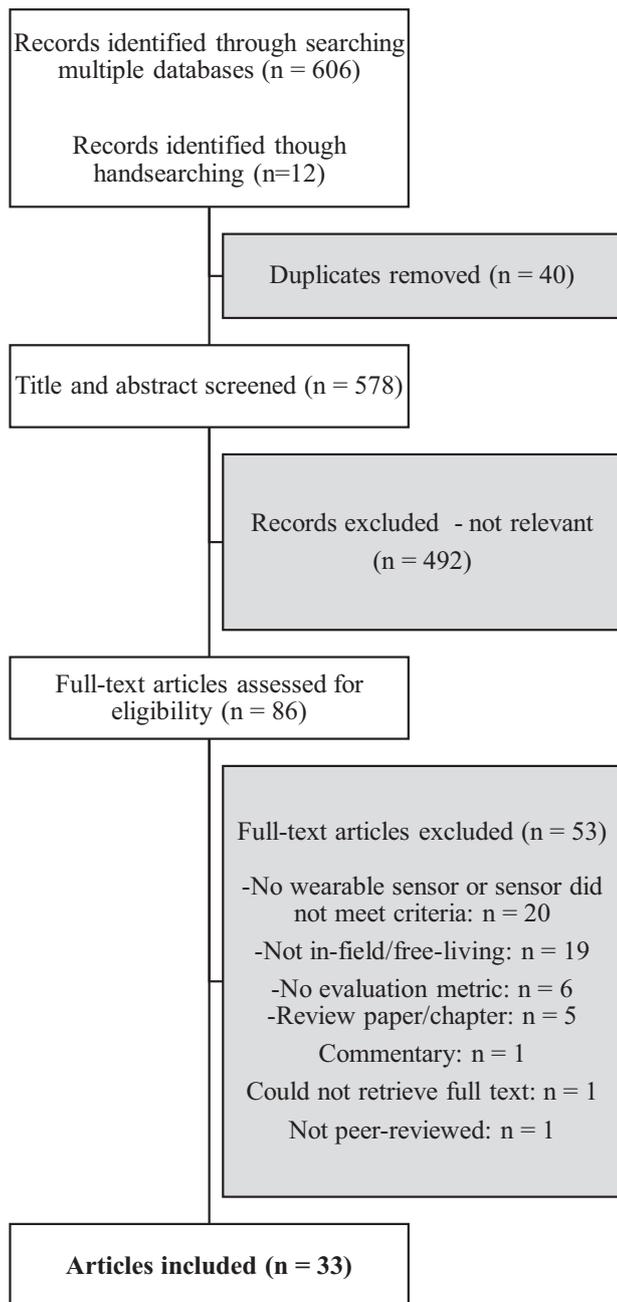
1. Any wearable device or sensor (i.e., worn on the body) that was used to automatically (i.e., no required actions by the user) detect any form of eating (e.g., content of food consumed, quantity of food consumed, eating event, etc.). Proxies for “eating” measures, such as glucose levels or energy expenditure, were not included. A “device” is defined as a wearable with embedded sensors that senses the body or environment and records data either on-device or externally through Bluetooth.
2. “In-field” (non-lab) testing of the sensor(s), in which eating and activities were performed at-will with no restrictions (i.e., what, where, with whom, when, and how the user ate could not be restricted). There was no limitation on the length of in-field testing. Furthermore, studies in which participants completed lab procedures (e.g., sensor calibration) followed by an in-field observation period that included eating were included. Studies that were described as semi-free living (e.g., participants wore device in the wild without restriction, but ate their meals in the lab), or studies that did not fit our aforementioned definition of “in-field”, were excluded.
3. At least one evaluation metric (e.g., Accuracy, Sensitivity, Precision, F1-score) that indicated the performance of the sensor on detecting its respective form of eating.

Review articles, commentary articles, study protocol articles, and any other articles without reported results from empirical research were excluded.

### Screening and selection of articles

The records retrieved from the databases were imported into a web-based systematic review management tool, Covidence<sup>32</sup>. Any duplicate articles were removed.

First, the titles and abstracts of all articles were reviewed. Articles were excluded at this initial screening phase if they did not describe at least one wearable device or sensor that was used to automatically detect any form of eating. If this information



**Fig. 1** Flow diagram of article selection process.

could not be ascertained from the title and/or abstract, the article's full text was reviewed in the next screening phase to determine whether it fit the eligibility criteria. Furthermore, if the abstract indicated that the study was not conducted in the field under free-living conditions, it was excluded. Otherwise, this criterion was assessed in the full-text screening phase (Fig. 1).

After the initial title/abstract screening process, the full texts of any remaining articles were assessed. Articles were further excluded if they did not meet the remaining eligibility criteria, including studies being conducted in free-living conditions; reporting at least one evaluation metric of the sensor's performance; being peer-reviewed; and reporting results from empirical research (Fig. 1).

#### Data extraction and synthesis

If any research article reported on more than one in-field study that fit our eligibility criteria (and contained different subject samples), then all eligible studies from that article were included in our review as separate studies.

Two research assistants independently extracted the following study characteristics from the final set of eligible studies using a custom-made data extraction worksheet. The extracted study characteristics were reviewed by first author BMB, and discrepancies were discussed and resolved by authors BMB and RA.

1. Study sample size: the total number of participants for each study.
2. Length of data collection period in free-living environment: how long the participants wore the sensor(s) (i.e., participated in the study) in the field.
3. Number of wearable sensors used to detect eating: the total number of wearable sensors in which the sensor's signal data were used, at least in part, to detect eating. Any other equipment that was part of the overall sensor system, but did not ultimately contribute signal data to detect eating, was not included in this count.
4. Types of wearable sensor(s) used to automatically detect eating: the type of sensors embedded within the selected wearable devices in which the sensor's signal data were used, at least in part, to detect eating.
5. Ground-truth method(s): the method, if any, that was used in the study to evaluate the performance of the sensor(s) in the free-living environment.
6. Form of eating activity (eating outcome) measured by the sensor(s): eating outcomes that were at least partially derived from sensor signal data and had a corresponding evaluation metric. In some cases, multiple eating outcomes for a single study were included. Terminology used to describe the eating outcome in the original article was maintained in our reporting.
7. Evaluation metric(s) of the sensor(s): any evaluation metric, reported either in text, a table, or a figure, that described the performance of the wearable sensor(s) on detecting eating in free-living settings. If the exact numerical value was not reported or could not be ascertained from a figure, and it was not reported numerically anywhere else in the text, then this evaluation metric was not included in Table 1. Only evaluation metrics that were exclusively for eating activity detection were included. Evaluation metrics that included the detection of other activities (e.g., sleeping, exercising, etc.) were not included. Often, multiple metrics were reported for different methods/algorithms deployed. If the article's authors indicated or highlighted the best performing method or algorithm result, then this was selected and included in the table for each unique method. Otherwise, the best performing (typically the highest value) of each unique method was identified in the article by the first author (BMB) and included in Table 1. Last, we exclusively included the metrics reported in the studies, and did not calculate any metrics, even if calculations were possible. For example, if Sensitivity and Specificity were reported, but Accuracy was not reported, we did not calculate Accuracy even though possible with the reported data.

#### Methods of summarizing data

Mean and standard deviation were calculated for each extracted continuous variable (sample size, length of time for data collection, number of sensor types, evaluation metric). Frequency tables were constructed for each extracted categorical variable (sensor types, ground-truth methods, eating outcomes).

Furthermore, the reported eating outcomes were grouped into three categories:

- i. inferred eating occasions: the incident or event of eating activity;
- ii. inferred chews (mastication): the grinding of food into smaller pieces by teeth;
- iii. inferred hand-to-mouth gestures: the placement of food into the mouth via movement of the hand to the mouth (also referred to as "bite").

**Table 1.** Methods and performance reported in wearable-based eating detection research studies ( $N = 40$ ).

First author, Year	Sample size	Free-living session duration	No. of sensor types	Sensor type(s)	Ground-truth method(s)	Eating outcome(s)	Select evaluation metric(s) <sup>a</sup>
Bedri, 2015 <sup>33</sup>	6	6 h	2	(1) Gyroscope (2) Infrared (IR) proximity sensor	Activity log	Eating events	Accuracy = 82.3% Avg. false positives/h: 1.7 Precision = 41.7% Recall = 90.1%
Bedri, 2017 <sup>34</sup>	10	3 h (x2 days)	3	(1) Inertial measurement unit (IMU) <sup>b</sup> (2) IR proximity sensor (3) Microphone	Wearable video camera (1-s resolution)	Chewing	Accuracy = 93% F1-score = 80.1% Precision = 81.2% Recall = 79%
Bi, 2018 <sup>35</sup>	14	2 h	1	(1) Microphone	Wearable video camera (1-s resolution)	Eating episodes	False positives: 2 True Positives: 15 (of 16) <i>Jaccard similarity coefficient</i> : Correctly detected episodes: 20 (of 26) Falsely detected episodes: 12 Missed episodes: 6 (of 26) <i>Ward's metrics</i> : Correctly detected episodes: 24 (of 26) Deletions: 2 (of 26) False insertions: 12
Blechert, 2017 <sup>36</sup>	14	1 day and 1 night	1	(1) Electromyogram (EMG) electrodes	Eating log via smartphone app; marker button on recording device	Eating episodes	Sensitivity = 87.3% Specificity = 86.9%
Chun, 2018 <sup>37</sup>	15	1 day	2	(1) Accelerometer (in smartphone) (2) IR proximity sensor	Eating log via smartphone app	Eating episodes	Precision = 78.2% Recall = 72.5%
Dong, 2011 <sup>38</sup>	4	1 day	3	(1) Accelerometer (2) Gyroscope (3) Magnetometer	Activity log	Eating activity	False negatives: 3 False positives: 6 Positive predictive value = 70% Sensitivity = 82%
Dong, 2014 <sup>39</sup>	43	1 day	2	(1) Accelerometer (in smartphone) (2) Gyroscope (in smartphone)	Eating log (first 20 participants); eating log via smartphone marker button (remaining 23 participants)	Meals/snacks	Accuracy = 81% Sensitivity = 81% Specificity = 82%
Doulah, 2017 <sup>18</sup>	8	24 h	3	(1) Accelerometer (2) Piezoelectric strain gauge sensor <sup>c</sup> (3) Radio-frequency (RF) transmitter and receiver	Eating log; push button	Eating episode duration	<i>Bland-Altman analysis</i> : Good agreement between automatic ingestion monitor & push button, but poor agreement between eating log and other methods. <i>Multiple comparisons analysis</i> : Eating episode durations from eating log were sig. diff. from automatic ingestion monitor ( $p < 0.001$ ) and push button ( $p < 0.001$ ). <i>Artificial Neural Network technique</i> : Accuracy = 86.86% Precision = 87.59% Sensitivity = 86.13%
Farooq, 2013 <sup>40</sup>	12	24 h	3	(1) Accelerometer (2) Piezoelectric strain gauge sensor <sup>c</sup>	Activity log; push button	Food intake	

Table 1 continued

First author, Year	Sample size	Free-living session duration	No. of sensor types	Sensor type(s)	Ground-truth method(s)	Eating outcome(s)	Select evaluation metric(s) <sup>a</sup>
1	~48 h	3	(3) RF transmitter and receiver (1) Accelerometer (2) Piezoelectric strain gauge sensor <sup>c</sup> (3) RF transmitter and receiver	Activity log	Meal episodes	Support Vector Machine technique: Accuracy = 81.93% Precision = 83.76% Sensitivity = 80.10% Artificial Neural Network technique: Correctly identified episodes: 8 (of 11) False positives: 1 Support Vector Machine technique: Correctly identified episodes: 7 (of 11) False positives: 3	
Farooq, 2016 <sup>41</sup>	24 h	3	(1) Accelerometer (2) Piezoelectric film sensor <sup>c</sup> (3) RF transmitter and receiver	Activity log; push button	Food intake	Fisher's Linear Discriminate Analysis (Bagging technique): Accuracy = 93.11% Precision = 96.72% Recall = 89.51%	
Farooq, 2017 <sup>42</sup>	≤3 h	1	(1) Piezoelectric film sensor	Portable tally counter; push button	Chews	Chew count estimation error (mean absolute value) = 6.24%	
Farooq, 2018 <sup>43</sup>	≤3 h	1	(1) Accelerometer	Eating log; push button	Food intake	F1-score = 85.8% Precision = 88.6% Recall = 85.4%	
Fontana, 2013 <sup>44</sup>	24 h	3	(1) Piezoelectric strain gauge sensor <sup>c</sup> (2) RF transmitter and receiver	Push button	Food intake	Accuracy = 73.2%	
Fontana, 2014 <sup>45</sup>	24 h	3	(1) Accelerometer (2) Piezoelectric film sensor (3) RF transmitter and receiver	Eating log; push button	Food intake	Using piezoelectric sensor and accelerometer data only: Accuracy = 89.8% Precision = 89.8% Recall = 89.9%	
Fortuna, 2016 <sup>46</sup>	Several hours	2	(1) Accelerometer (2) Gyroscope	Wearable video camera (10-s resolution)	Hand-to-mouth motions ("bites")	Subjects 1 and 2: Accuracy = 96.9% Precision = ~70% Recall = ~70% Subject 3: Accuracy = ~82% Precision = ~18% Recall = ~70%	
Gao, 2016 <sup>47</sup>	More than 10 days	1	(1) Microphone (in Bluetooth headset)	Participants recorded their eating episodes with a smartphone front-facing video camera	Eating episodes	Leave One Sample Out (LOSO) method: Accuracy = 75.61% Deep learning accuracy = 94.72% Leave One Person Out (LOPO) method: Accuracy = 65.43% Deep learning accuracy = 76.82%	
Gomes, 2019 <sup>48</sup>	3–5 h	2	(1) Accelerometer (in IMU) (2) Gyroscope (in IMU)	Real-time annotation via mobile phone prompt	Hand-to-mouth movements preceding a drinking event	Deep learning accuracy = 76.82% False negatives = 24 False positives = 27 True positives = 113	

Table 1 continued

First author, Year	Sample size	Free-living session duration	No. of sensor types	Sensor type(s)	Ground-truth method(s)	Eating outcome(s)	Select evaluation metric(s) <sup>a</sup>
Hamatani, 2018 <sup>49</sup>	16	1 day	2	(1) Accelerometer (in smartwatch) (2) Gyroscope (in smartwatch)	Wearable video camera (5-s resolution)	Drinking activity	F-score = 0.85 Precision = 0.84 Recall = 0.85 False negatives: 40 False positives: 31 True positives: 138 Precision = 81.7% Recall = 77.5% Mean absolute percentage error = 31.8% Mean percentage error = 14.0% Overall error = 4.3%
	8	2 days	2	(1) Accelerometer (in smartwatch) (2) Gyroscope (in smartwatch)	Wearable video camera (5-s resolution)	Drinking activity	False negatives: 33 False positives: 41 True positives: 81 Precision = 66.4% Recall = 71.1% Mean absolute percentage error = 34.6% Mean percentage error = 6.9% Overall error = -15.9%
Jia, 2018 <sup>50</sup>	1	1 week	1	(1) Wearable video camera	Wearable video camera (10-s resolution)	Food and drink images (episodes)	When Burden Index = 18% (k = 2): Sensitivity = 74.0% Specificity = 87.0%
Kyritsis, 2019 <sup>51</sup>	6	Not reported (on average, 17,398 s per person)	2	(1) Accelerometer (in smartwatch) (2) Gyroscope (in smartwatch)	Eating log	Meals	False negatives = 55,083 (s) False positives = 47,187 (s) True negatives = 6,424,247 (s) True positives = 432,917 (s) Average Jaccard Index = 0.804 F1-score = 0.894 Precision = 0.901 Recall = 0.887 Specificity = 0.992
Mirtchouk, 2017 <sup>52</sup>	5	12 h (x2 days)	3	(1) Inertial motion sensor (in smartwatch) (2) Microphone (3) Motion sensor (in Google Glass)	Eating log; photos of meals; voice notes	Meals	Accuracy = 85% Precision = 31% Recall = 87%
	6	2 or 5 days	2	(1) Inertial motion sensor (in smartwatch) (2) Microphone	Eating log; photos of meals; voice notes	Meals	Accuracy = 79% Precision = 25% Recall = 83%
Navarathna, 2018 <sup>53</sup>	1	1 month	4	(1) Accelerometer (in smartphone) (2) Accelerometer (in smartwatch) (3) GPS (in smartphone) (4) Gyroscope (in smartwatch)	Activity log via smartphone app	Eating activity	Using gyroscope data only: Correctly predicted eating activity: 1206

Table 1 continued

First author, Year	Sample size	Free-living session duration	No. of sensor types	Sensor type(s)	Ground-truth method(s)	Eating outcome(s)	Select evaluation metric(s) <sup>a</sup>
Rahman, 2016 <sup>54</sup>	8	5 days	4	(1) Affectiva Q Sensor <sup>d</sup> (2) GPS (in smartphone) (3) Microphone (4) Microsoft Band <sup>e</sup>	Eating log via smartphone app	"About-to-Eat" moments	F-score = 0.69 Precision = 0.67 Recall = 0.77
Schiboni, 2018 <sup>55</sup>	1	4 days	1	(1) Wearable video camera	Wearable video camera (1-s resolution)	Dietary events	F1-score = 90% Mean average precision (mAP) = 51% Precision = 78% Recall = 100%
Sen, 2017 <sup>56</sup>	7	5 days	3	(1) Accelerometer (in smartwatch) (2) Camera (in smartwatch) (3) Gyroscope (in smartwatch)	Smartwatch camera (participants validated the images via smartphone food journal at end of the day)	Eating periods	False negatives = 0% False positives = 60.3% True positives: 31
	6	2 days	3	(1) Accelerometer (in smartwatch) (2) Camera (in smartwatch) (3) Gyroscope (in smartwatch)	Smartwatch camera (participants validated the images via smartphone food journal at end of the day)	Eating periods	False negatives = 35.3% False positives = 31.3% True positives: 11
	4	5 days	3	(1) Accelerometer (in smartwatch) (2) Camera (in smartwatch) (3) Gyroscope (in smartwatch)	Smartwatch camera (participants validated the images via smartphone food journal at end of the day)	Eating periods	False negatives = 3.3% False positives = 23.7% True positives: 29
Sen, 2018 <sup>57</sup>	4	5 days	3	(1) Accelerometer (in smartwatch) (2) Camera (in smartwatch) (3) Gyroscope (in smartwatch)	Smartwatch camera (participants validated the images via smartphone food journal at end of the day)	Eating episodes	False negatives: 1 (of 30) False positives: 2 True positives: 29 (of 30)
	5	4–6 days	3	(1) Accelerometer (in smartwatch) (2) Camera (in smartwatch) (3) Gyroscope (in smartwatch)	Smartwatch camera (participants validated the images via smartphone food journal at end of the day)	Meals	False negatives: 3 (of 51) False positives: 2 (of 51) Combined (study 1 + 2): Precision = 95% (after image filtering) Recall = 95%
Sharma, 2016 <sup>58</sup>	104	1 day	2	(1) Accelerometer (2) Gyroscope	Eating log	Eating activity	Accuracy = 75% Sensitivity = 69% Specificity = 80%
Thomaz, 2015a <sup>59</sup>	7	Not reported (on average, 5 h 42 minutes per person)	1	(1) Accelerometer (in smartwatch)	Wearable video camera (60-s resolution)	Eating moments	F-score = 76.1% Precision = 66.7% Recall = 88.8%
	1	31 days	1	(1) Accelerometer (in smartwatch)	Wearable video camera (60-s resolution)	Eating moments	

**Table 1** continued

First author, Year	Sample size	Free-living session duration	No. of sensor types	Sensor type(s)	Ground-truth method(s)	Eating outcome(s)	Select evaluation metric(s) <sup>a</sup>
Thomaz, 2015 <sup>b</sup>	20	4–7 h	1	(1) Microphone	Activity log	Meal eating activity	F-score = 71.3% Precision = 65.2% Recall = 78.6% F-score = 79.8%; Precision = 89.6%; Recall = 76.3%;
Yatani, 2012 <sup>c1</sup>	5	1 day	1	(1) Microphone	Wearable video camera (in smartphone) (resolution not reported)	Eating activity	<i>Support Vector Machine classification:</i> Correctly predicted eating activities: 157 Precision = 81.3% Recall = 87.8% <i>Naive Bayes classification:</i> Correctly predicted eating activities: 125 Precision = 62.2% Recall = 69.8%
Ye, 2016 <sup>d2</sup>	7	2 weeks	1	(1) Accelerometer (in smartwatch)	Eating log in Evernote app; smartwatch physical button pushed to confirm/deny eating activity	Eating	<i>Support Vector Machine classification:</i> Correctly predicted drinking activities: 33 Precision = 61.1% Recall = 56.0% <i>Naive Bayes classification:</i> Correctly predicted drinking activities: 36 Precision = 28.6% Recall = 61.0%
Zhang, 2018a <sup>e3</sup>	10	1 day	1	(1) EMG electrodes	Eating log	Eating events	False detections per subject per day: ~7 Precision = ~31%
Zhang, 2018b <sup>e4</sup>	10	1 day	1	(1) EMG electrodes	Activity log	Chewing	F1-score = 95.2% Precision = 98.2% Recall = 98.7%
						Eating events	Precision = 77.3% Recall = 78.8%
							False negatives: 1 (of 44) False positives: 0 True positives: 43 (of 44) Precision > 0.95 Recall > 0.95

<sup>a</sup>The highest reported value for each unique evaluation metric is reported in this table. If multiple methods and/or algorithms were evaluated within a single study, the highest reported value for each unique evaluation metric for each method is reported.

<sup>b</sup>An inertial measurement unit (IMU) is a device that is typically comprised of a combination of accelerometers, gyroscopes, and sometimes magnetometers.

<sup>c</sup>Additional information needed to adequately describe this sensor was extracted from a paper previously published by the same author, in which the sensor is described in more detail (Sazonov and Fontana<sup>83</sup>).

<sup>d</sup>The Affectiva Q sensor measures electrodermal activity.

<sup>e</sup>The Microsoft Band is a smartwatch/fitness tracker that contains many sensors, including an accelerometer, gyroscope, heart rate monitor, and skin temperature sensor.

These three categories are not exhaustive of all possible types of eating outcome categories, but rather only represent those that were derived from this review's included studies.

## RESULTS

### Literature search

The literature search produced 618 research articles, with 40 duplicates, resulting in 578 articles to be screened. After reviewing all article titles and abstracts, the full texts of 86 of these articles were further reviewed for eligibility. After removing articles that did not meet the inclusion criteria (see Fig. 1 for full list of exclusion reasons), 33 articles were deemed eligible for the review<sup>18,33–64</sup> (Fig. 1). Six of these articles<sup>40,49,52,56,57,59</sup> reported on more than one in-field study that fit our eligibility criteria, so  $N = 40$  studies (from 33 articles) is considered to be the final sample size for the review.

### Study characteristics

The earliest publication year of the reviewed papers was 2011, and the most recent year was 2019. The sample size of the studies ranged from 1 to 104 participants, with a mean of 10.83 participants ( $SD = 16.73$ ) per study (Table 1). The length of time for data collection in the “wild” environment varied and was not always reported with an exact numerical value or unit. Therefore, we will just report the range: 2 h to 1 month. The reported length of time for each study is available in Table 1.

### Wearable sensors

The majority of studies ( $N = 26$  of 40) used multi-sensor systems (incorporating  $> 1$  wearable sensor) to automatically detect eating activity<sup>18,33,34,37–41,44–46,48,49,51–54,56–58</sup>. On average, 2.10 wearable sensors ( $SD = 0.96$ ) were used in the studies, with a range of 1–4 sensors (Table 1). Approximately 63% ( $N = 25$ ) of the 40 studies utilized an accelerometer (device that determines acceleration) either by itself ( $N = 4$ ) or incorporated into a sensor system ( $N = 21$ ) to detect eating activity<sup>18,37–41,43,45,46,48,49,51,53,56–59,62</sup> (Table 2). The second most frequently utilized wearable sensor was a gyroscope (device that determines orientation) ( $N = 15$ )<sup>33,38,39,46,48,49,51,53,56–58</sup>, followed by a microphone ( $N =$

8)<sup>34,35,47,52,54,60,61</sup>, a piezoelectric sensor ( $N = 7$ )<sup>18,40–42,44,45</sup>, a RF transmitter and receiver ( $N = 6$ )<sup>18,40,41,44,45</sup>, and a smartwatch camera ( $N = 5$ )<sup>56,57</sup> (Table 2). EMG electrodes<sup>36,63,64</sup>, a motion sensor<sup>52</sup>, and an infrared proximity sensor<sup>33,34,37</sup> were used in three studies each. Wearable video cameras<sup>50,55</sup> and GPS<sup>53,54</sup> were used in two studies. Last, the Affectiva Q sensor (used to measure electrodermal activity)<sup>54</sup>, an inertial measurement unit (IMU)<sup>34</sup>, a magnetometer<sup>38</sup>, and a Microsoft Band<sup>54</sup> were used in one study each (Table 2).

### Ground-truth methods

To evaluate the performance of the wearable sensors in automatically detecting eating in the field, 100% of studies ( $N = 40$ ) used one or more validation methods in which the “ground-truth” eating data reported by the participants (self-report methods) or produced by other wearable devices (objective methods) were compared with the inferred eating activity detected by the sensor(s).

**Self-report methods.** Six of the studies had participants self-report all daily activities, including eating, via a log or diary<sup>33,38,40,53,60,64</sup>, while six studies had participants self-report just eating activity<sup>37,39,51,54,58,63</sup>. In one study, participants were asked to record their eating episodes with a smartphone front-facing video camera to obtain ground-truth eating<sup>47</sup>. In another study, participants used a push button, located on a wireless device, as the primary method for self-reporting food intake; participants pressed and held a button during chewing to indicate the start and end of a chewing bout<sup>44</sup>. In Farooq and Sazonov<sup>42</sup>, participants used a push button, and they also counted their number of chews during each chewing bout by using a portable tally counter (a device used to incrementally count something).

Other studies used multiple self-report methods to collect ground-truth eating activity. Six studies used both an eating or activity log in addition to a marker/push button<sup>18,36,40,41,43,45</sup>. Two studies, both reported in Mirtchouk et al.<sup>52</sup>, had participants use a combination of eating logs, voice notes (“spoken annotations”), and self-taken photos at the start and end of meals.

**Objective methods.** One-quarter of studies ( $N = 10$  of 40)<sup>34,35,46,49,50,55,59,61</sup> outfitted participants with a wearable video camera for the length of their respective data collection period, in which the video clips were later annotated by a person or persons from the research team to indicate periods of eating activity. The video cameras captured images at varying levels of resolution: three studies captured images in 1-s intervals; two studies captured images every 5 s; two studies captured images every 10 s; and two studies captured images every 60 s.

**Mixed methods.** Two studies sent participants real-time messages to confirm whether or not they were eating. In Ye et al.<sup>62</sup>, participants were sent a short message on their smartwatch if an eating gesture was detected; participants were able to confirm or reject if they were eating in real-time. Similarly, in Gomes and Sousa<sup>48</sup>, when drinking activity was detected, participants were sent an alert on their smartphone and could then confirm or reject if they were drinking in real-time.

Last, five studies (reported in two papers) used an “automated food journaling system”, which consisted of a combination of self-report (food journal), wearable sensor (smartwatch inertial sensors), and objective data (smartwatch camera) to discern ground-truth of eating activity in real-time<sup>56,57</sup>. When the wrist-worn smartwatch detected an eating episode, the smartwatch then captured images and sent them to a server. After image processing techniques were applied to eliminate irrelevant images, a subset of relevant images were stored in the server

**Table 2.** Frequency and percentage of sensor types in included studies, ordered by frequency.

Sensor type	Frequency	Percentage (of 40 studies)
Accelerometer	25	62.50%
Gyroscope	15	37.50%
Microphone	8	20.00%
Piezoelectric sensor	7	17.50%
Radio-frequency transmitter and receiver	6	15.00%
Smartwatch camera	5	12.50%
Electromyogram electrodes	3	7.50%
Motion sensor	3	7.50%
Infrared proximity sensor	3	7.50%
GPS	2	5.00%
Wearable video camera	2	5.00%
Affectiva Q Sensor	1	2.50%
Inertial measurement unit	1	2.50%
Magnetometer	1	2.50%
Microsoft Band	1	2.50%

**Table 3.** Reported eating outcomes ( $n = 45$ ) from included studies, by category.

Category 1: Eating occasions ( $n = 40$ )	Category 2: Chews ( $n = 3$ )	Category 3: Hand-to-mouth gestures ( $n = 2$ )
"About-to-Eat" moments: 1	Chewing: 2	Hand-to-mouth motions: 1
Dietary events: 1	Chews: 1	Hand-to-mouth movements preceding a drinking event: 1
Drinking activity: 3		
Eating: 1		
Eating activity: 4		
Eating episode duration: 1		
Eating episodes: 6		
Eating events: 3		
Eating moments: 2		
Eating periods: 3		
Fluid intake: 2		
Food & drink images (episodes): 1		
Food intake: 5		
Meal eating activity: 1		
Meal episodes: 1		
Meals: 4		
Meals/snacks: 1		

and could be viewed by the user to assist with their self-report online food journaling<sup>56,57</sup>.

### Eating outcomes

Authors collectively reported on 22 different types of eating outcomes (Table 1). For each study, only the eating outcome that was (a) directly measured by the wearable sensors, and (b) evaluated with comparison to a ground-truth method is included in Tables 1 and 3 and is reported in this section. Five studies<sup>34,49,61,64</sup> reported two eating outcomes each that fit these criteria, therefore the five additional outcomes were included, totaling 45 eating outcomes.

The reported measures were grouped into three categories: (1) inferred eating occasions, (2) inferred chews, and (3) inferred hand-to-mouth gestures (Table 3). The majority of studies ( $N = 37$  of 40) used wearable sensors to infer eating occasions (Category 1), however, these measures were often referred to in different terms (e.g., "eating event" vs. "eating episode" vs. "dietary event"). Three studies inferred chews (Category 2)<sup>34,42,64</sup> and two studies inferred hand-to-mouth gestures (Category 3)<sup>46,48</sup>.

### Sensor evaluation metrics

Studies often reported multiple and varied evaluation metrics. All reported evaluation metrics for their corresponding eating outcome(s) are included in Table 1. The most frequently reported metrics were Accuracy and F1-score. Twelve studies reported Accuracy<sup>33,34,39–41,44–47,52,58</sup> and 10 studies reported F1-score<sup>34,43,48,51,54,55,59,60,63</sup>. Due to the lack of a standardized evaluation metric across studies, we do not summarize (calculate mean, standard deviation, etc.) the reported metrics. However, select evaluation metric values for each study are available in Table 1.

## DISCUSSION

Detecting and monitoring human eating behavior in the real-world is a growing field and offers exciting and novel opportunities for future public health research. Previous literature reviews on automatic eating detection methods have often placed a focus on sensor modalities and technologies in any research setting<sup>28</sup>, on specific sensor modalities, such as wearable video cameras<sup>65</sup>, or on specific sensor locations, such as the upper limb<sup>29</sup>. No review to date has focused specifically on studies that use wearable devices to automatically detect eating activity in naturalistic settings. Furthermore, this review has been written by a multidisciplinary team, with a multidisciplinary audience in mind. Below we provide a discussion on trends in wearable sensors for eating detection, a summary of technical, analytical, and multidisciplinary challenges that occur when conducting in-the-wild research, and recommendations for future public health research directions.

### Summary of key findings

This scoping review included 40 studies that reported on the use of wearable sensors to automatically detect eating in the field published through December 2019. The majority of studies utilized accelerometers ( $N = 25$  of 40) and/or gyroscopes ( $N = 15$  of 40) to automatically detect eating activity via wearable sensors. Self-reported eating was the most frequently used method for collecting ground-truth eating activity data, whereas objective methods such as wearable video cameras were used in only one-quarter of studies ( $N = 10$  of 40).

Variations of 'inferred eating occasions' (e.g., eating episode, eating event, eating period) were the most commonly reported eating measure ( $N = 37$ ), followed by inferred chews ( $N = 3$ ), and then inferred hand-to-mouth gestures ( $N = 2$ ). The most commonly reported evaluation metrics were Accuracy ( $N = 12$ ) and F1-score ( $N = 10$ ).

In reviewing the methods and performance of wearable devices and sensors that automatically detect eating in naturalistic research settings, several trends and major challenges can be identified.

### Trends, challenges, and recommendations

**Sensors.** In this review, we observe that early research efforts attempted to find novelty and improvement by experimenting with new sensor types and/or sensor locations. The first paper to report the use of a single sensor to detect eating was in 2012<sup>61</sup>, but it did not become more prevalent until 2015<sup>59,60</sup>; it became much more popular by 2018, in which more than half of the 10 papers published that year only used a single sensor<sup>35,43,50,55,63,64</sup>. With time, these efforts have converged into a shorter list of sensor types and locations with emphasis on two major criteria: the sensors' ability to capture the patterns of eating outcomes, and the real-world practicality, which includes user comfort, acceptability, and accessibility of continuously wearing these sensors in a real-world setting for a long duration of time. As seen in Tables 1 and 2, the majority of studies utilized accelerometers ( $N = 25$ ) and/or gyroscopes ( $N = 15$ ), typically embedded within a wrist-worn smartwatch or device. This observation highlights the emergence of smartwatches as a potentially practical and user-friendly modality for real-world eating behavior monitoring. In addition to the user advantages, smartwatch-based methods also have engineering advantages, such as facilitating compact and concise form factor, containing reasonable computational and power resources (e.g., low battery footprint), and offering wide possibilities for adoption with the rising Internet of Things (IoT)-cloud-edge-based ubiquitous computing era.

Another noted trend is the combination of multiple sensors' information from different modalities placed on various on-body

locations (for example, the piezoelectric strain gauge-infrared proximity sensor-accelerometer combination used in a few studies<sup>18,40,41,44</sup>). However, using a 'multi-sensor system' (i.e., multiple sensors) as opposed to a single sensor introduces a multitude of challenges, ranging from having to integrate different sampling rates and signal amplitudes to issues arising from data synchronization (i.e., aligning sensor signals in time across multiple devices due to differences in clock times) and reliability (i.e., ensuring a consistent sampling rate to maintain high quality data throughout the study). Despite these challenges, this multi-sensor approach offers high potential for real-time monitoring and tracking of human eating behavior. Multi-sensor fusion can provide further context-awareness, which can aid in modeling of behavior; and in optimization of power (e.g., a low power sensor can trigger a higher power sensor only when necessary) and computational complexities, which is vital for sensor-driven monitoring in real-world settings. Furthermore, it can provide new opportunities for simultaneously detecting eating behavior (e.g., when people are eating), dietary intake (e.g., what people are eating), and context (e.g., where people are eating, with whom), culminating in a much fuller understanding of individual eating patterns and dynamics.

Two papers that were retrieved in the original literature search, but were not ultimately included in the review because they did not report an evaluation metric of the automatic eating detection performance, have started to do this. Vaizman et al. and Gemming et al. report on using wearable sensors and mobile devices to automatically capture both eating and contexts in natural environments (free living)<sup>66,67</sup>. In Gemming et al.<sup>66</sup>, social and contexts such as eating location, external environment (indoor/outdoor), physical position, social interaction, and viewing media screens are collected. Similarly, Vaizman et al.<sup>67</sup> collect contexts such as location (e.g., in class, in a meeting, at a restaurant), activities (e.g., running, drinking (alcohol), eating), and company (e.g., with friends). These recent works suggest that automatic eating and context detection methods offer opportunities to provide valuable, rich, real-time information with important obesity-prevention research implications.

**Ground-truth methods.** Our review indicates that there is still a strong reliance on self-report as the method of determining ground-truth eating activity in the field. Although using objective methods such as wearable video cameras<sup>68</sup> to determine when eating activity has occurred can be very costly and time-intensive, maintaining eating and/or activity logs is a process that is very burdensome on the participant and ultimately relies on their memory on when food was eaten. In studies comparing food diaries to the doubly labeled water method, which uses doubly labeled water as a biomarker for energy intake and is considered the "gold-standard" method<sup>69</sup> of dietary assessment, there is evidence that misreporting of dietary intake is common<sup>17</sup>. Moreover, these "gold-standard" methods have recently been shown to exhibit methodological biases<sup>70</sup>, particularly when individuals consume low-carbohydrate diets; and have only been used to evaluate self-reported dietary intake, not self-reported timing of eating activity. The accuracy of self-reported timing of eating activity via food diaries is unknown. The aforementioned limitations in current eating assessment methodology were often cited in the reviewed articles as the motivating factor for developing wearable technologies that can overcome these constraints. Therefore, the use of these self-report methods to validate the wearable sensors may need to be reconsidered in the future when evaluating the reported performance of these sensors.

The trend in ground-truth validation is increasingly in favor of using a wearable video camera because of the increased confidence established from visually confirming the activity or behavior being performed; 10 of the included studies

used wearable video cameras as ground-truth methods<sup>34,35,46,49,50,55,59,61</sup>, with six of those studies taking place in the last three years (2017–2019). However, privacy and stigma challenges associated with wearable video cameras remain when studying participants in free-living populations, particularly if the objective is to capture authentic eating behavior<sup>71</sup>. Several device- (i.e., video camera lens orientation, location, look and feel) and data-specific factors (i.e., what is being collected by the video camera) introduces discomfort, thereby influencing an individual's willingness to act naturally and results in wearers modifying their behavior or abandoning the device<sup>71</sup>. However, these devices may only need to be worn for a limited time, to aid in the process of validating and building machine learning-based models that detect eating in the real world.

Mobile device-assisted ecological momentary assessment (mEMA), repeated sampling of one's behavior in real-time and in-context<sup>24</sup>, has been suggested as a promising tool for eating assessment<sup>72</sup>, and may also serve as a novel tool to evaluate the performance of wearable sensors in the field. In a recent review, Schembre and colleagues<sup>73</sup> summarize the existing literature on mEMA methods for the measurement of dietary intake in research studies. They conclude that mobile ecological momentary dietary assessment methods may be interchangeable with existing methods, in addition to reducing participant burden, recall biases, and technological barriers while maximizing ecological validity<sup>73</sup>.

Two studies (Ye et al.<sup>62</sup>, Gomes and Sousa<sup>48</sup>) included in this review used a novel method for obtaining ground-truth eating activity in the wild similar to mEMA. When an eating/drinking gesture was detected via the wearable sensors, the watch or phone displayed a short message asking the user to confirm or reject that the user was eating/drinking. Similarly, in a study not included in this review, Spruijt-Metz et al.<sup>74</sup> report on an integrated system of wearable sensors and smartphones that collects family eating behavior data in the wild, and propose using mEMA to collect ground-truth eating data<sup>75</sup>. mEMA and similar methods do have their own limitations, such as relying on self-reported eating activity and the potential for low compliance, yet may offer the ability to capture and validate ground-truth eating activity in naturalistic settings for longer periods of time as compared with other subjective and objective measures, thus potentially improving research scalability and participant acceptability.

Exploration and implementation of new methods for acquiring ground-truth eating activity in naturalistic settings, as well as continued testing of the emerging methods such as wearable video cameras and mEMA, is warranted in future research in order to improve the validity and reliability of eating detection via wearable sensors.

#### Limitations of the scoping review

To our knowledge, this is one of the first review papers to date that catalogs and summarizes the current use of wearable sensors that automatically detect eating activity specifically in free-living or uncontrolled research settings. However, due to this strict inclusion criteria, we did not include papers that reported on sensors deployed in semi-controlled settings, which may have contributed to a more comprehensive review of the current sensors being used in this field of research. In addition, the lack of both a standardized eating outcome (see Table 3) and a standardized evaluation measure of said eating activity prevented us from comparing successes across studies. Moreover, we selected and reported the 'best' evaluation metric from studies in Table 1 (typically the highest value), thus potentially biasing the overall takeaway message of how accurate these sensors are at detecting eating in the wild. Last, some evaluation metric results for the included papers were exclusively reported in figures (not

textually or numerically), and were omitted from this review due to the data extraction protocol.

#### From research to practice

Recent technological advances have provided engineers with the capability to develop sensors and devices that can collect the data needed to objectively infer eating activity in a variety of novel ways, such as inferring bites via data from wrist-worn sensors<sup>46</sup>, sensing jaw movement to infer chewing activity<sup>42</sup>, and utilizing the signal obtained from EMG electrodes to infer when eating events occur<sup>63</sup>. This variety of sensors and abundance of inferable eating outcomes, however, contributes to the difficulty of comparing performance across sensors and evaluation methods. Some form of comparability among similar modalities will be necessary in order to integrate wearable sensors into public health research. This includes developing some standardized form of comparability, such as standardized terminology and standardized evaluation metrics, or at least the sharing of algorithms that are used to infer various eating behaviors. Importantly, identification of meaningful eating outcomes and the development of the algorithms to detect these outcomes would benefit from being generated with input from relevant collaborators such as computer scientists and engineers, health behaviorists, and nutritionists.

Researchers within the public health field are increasingly interested in utilizing wearable sensors to assess eating activity. However, diet, which is one of the most important factors that contributes to overweight and obesity status and other diseases, is notoriously hard to assess with our current methods<sup>16</sup>. Although flawed, more traditional methods offer assurance in what type of eating outcome is being measured. For instance, 24-h dietary recalls typically obtain detailed information on all foods and beverages consumed on a given day, which can later be processed into meaningful behaviors (at least, meaningful as determined by the nutrition field) such as average total intake of kilocalories per day or an overall healthy eating index. As indicated by the numerous eating outcomes extracted from the papers (see Table 3), the same assurance cannot necessarily be granted for wearable sensors, especially considering the evidence that how an eating occasion is defined can significantly influence how eating patterns are characterized<sup>76</sup>. The wide variation in sensor modalities makes this difficult, but also offers us novel research questions that researchers could never consider before (e.g., How is an individual's number of bites taken throughout the day associated with their risk for developing obesity? Does an individual's chewing rate predict cardiovascular disease risk? Etc.).

#### Future directions

Eating is a highly dynamic behavior that occurs over time and in context. Research has repeatedly shown that contexts have important impacts on eating patterns and health outcomes, including social contexts like eating with family<sup>77–79</sup>, physical contexts like living in a rural food environment<sup>80</sup>, and psychological contexts like chronic stress levels<sup>81,82</sup>. Eating data from wearable sensors, in conjunction with other technologies that can detect these additional contextual features, can be used to eventually develop models of human eating behavior in real time and in context, which can be leveraged to develop adaptive, personalized interventions to predict and change obesity-related behavior, and ultimately, improve population health.

#### CONCLUSIONS

How do the results from all of these papers inform field work, and how do all of the various sensor modalities reviewed here, really work in the wild? The field is still nascent, and the accuracy of our sensors are still dependent on careful constraints and

assumptions. Collecting data in the wild is considerably more challenging than conducting research in controlled settings. In the wild, we lose continuous control of our sensors, participants tamper with the hardware or software, new operating systems bring down our software temporarily. Therefore, successful deployment requires really robust sensing and algorithms that have been tried repeatedly, rather than re-invented again and again. If the goal is to be able to detect eating behaviors in a fully unconstrained manner over a long period of time, novelty can no longer be valued over accuracy. We hope that this review will convince our communities to undertake the hard work and collaboration necessary to develop the kinds of reliable technologies that are required to conduct long-term real-world studies. Developing accurate, robust, automatic and wearable-based eating detection methods will radically reform scientific understanding of eating behavior in time and context and public health.

Received: 1 July 2019; Accepted: 13 February 2020;

Published online: 13 March 2020

#### REFERENCES

1. Neuhouser, M. L. The importance of healthy dietary patterns in chronic disease prevention. *Nutr. Res.* **70**, 3–6 (2019).
2. Jannasch, F., Kröger, J. & Schulze, M. B. Dietary patterns and type 2 diabetes: a systematic literature review and meta-analysis of prospective studies. *J. Nutr.* **147**, 1174–1182 (2017).
3. Higgs, S. & Thomas, J. Social influences on eating. *Curr. Opin. Behav. Sci.* **9**, 1–6 (2016).
4. Tourlouki, E., Matalas, A.-L. & Panagiotakos, D. B. Dietary habits and cardiovascular disease risk in middle-aged and elderly populations: a review of evidence. *Clin. Inter. Aging* **4**, 319–330 (2009).
5. Robinson, E., Thomas, J., Aveyard, P. & Higgs, S. What everyone else is eating: a systematic review and meta-analysis of the effect of informational eating norms on eating behavior. *J. Acad. Nutr. Dietetics* **114**, 414–429 (2014).
6. Reicks, M. et al. Influence of parenting practices on eating behaviors of early adolescents during independent eating occasions: implications for obesity prevention. *Nutrients* **7**, 8783–8801 (2015).
7. Hales, C. M., Fryar, C. D., Carroll, M. D., Freedman, D. S. & Ogden, C. L. Trends in obesity and severe obesity prevalence in us youth and adults by sex and age, 2007–2008 to 2015–2016. *JAMA* **319**, 1723–1725 (2018).
8. Fryar, C. D., Carroll, M. D. & Ogden, C. L. *Prevalence of Obesity among Children and Adolescents: United States, Trends 1963–1965 through 2009–2010* (National Center for Health Statistics, 2012).
9. Fryar, C. D., Carroll, M. D. & Ogden, C. L. *Prevalence of Overweight, Obesity, and Extreme Obesity among Adults: United States, Trends 1960–1962 through 2009–2010* (National Center for Health Statistics, 2012).
10. Afshin, A. et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **393**, 1958–1972 (2019).
11. Shim, J.-S., Oh, K. & Kim, H. C. Dietary assessment methods in epidemiologic studies. *Epidemiol. Health* **36**, e2014009 (2014).
12. Thompson, F. E., Subar, A. F., Loria, C. M., Reedy, J. L. & Baranowski, T. Need for technological innovation in dietary assessment. *J. Am. Dietetic Assoc.* **110**, 48–51 (2010).
13. Magarey, A. et al. Assessing dietary intake in children and adolescents: considerations and recommendations for obesity research. *Int. J. Pediatr. Obes.* **6**, 2–11 (2011).
14. Willett, W. in *Nutritional Epidemiology* (Oxford University Press, 1998).
15. Livingstone, M. B., Robson, P. J. & Wallace, J. M. Issues in dietary intake assessment of children and adolescents. *Br. J. Nutr.* **92**(Suppl 2), S213–S222 (2004).
16. Westertep, K. R. & Goris, A. H. C. Validity of the assessment of dietary intake: problems of misreporting. *Curr. Opin. Clin. Nutr. Metab. Care* **5**, 489–493 (2002).
17. Schoeller, D. A. Limitations in the assessment of dietary energy intake by self-report. *Metabolism* **44**, 18–22 (1995).
18. Doulah, A. et al. Meal microstructure characterization from sensor-based food intake detection. *Front. Nutr.* **4**, 31 (2017).
19. Marcum, C. S., Goldring, M. R., McBride, C. M. & Persky, S. Modeling dynamic food choice processes to understand dietary intervention effects. *Ann. Behav. Med.* **52**, 252–261 (2018).

20. Ohkuma, T. et al. Association between eating rate and obesity: a systematic review and meta-analysis. *Int J. Obes.* **39**, 1589–1596 (2015).
21. Bell, B. M. et al. Sensing eating mimicry among family members. *Transl. Behav. Med.* **9**, 422–430 (2019).
22. Sharps, M. et al. Examining evidence for behavioural mimicry of parental eating by adolescent females. An observational study. *Appetite* **89**, 56–61 (2015).
23. Subar, A. F. et al. The Automated Self-Administered 24-hour dietary recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. *J. Acad. Nutr. Diet.* **112**, 1134–1137 (2012).
24. Shiffman, S., Stone, A. A. & Hufford, M. R. Ecological momentary assessment. *Annu Rev. Clin. Psychol.* **4**, 1–32 (2008).
25. Boushey, C. J., Spoden, M., Zhu, F. M., Delp, E. J. & Kerr, D. A. New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. *Proc. Nutr. Soc.* **76**, 283–294 (2017).
26. Spruijt-Metz, D. et al. Advances and controversies in diet and physical activity measurement in youth. *Am. J. Preventive Med.* **55**, e81–e91 (2018).
27. McClung, H. L. et al. Dietary intake and physical activity assessment: current tools, techniques, and technologies for use in adult populations. *Am. J. Preventive Med.* **55**, e93–e104 (2018).
28. Vu, T., Lin, F., Alshurafa, N. & Xu, W. Wearable food intake monitoring technologies: a comprehensive review. *Computers* **6**, 4 (2017).
29. Heydarian, H., Adam, M., Burrows, T., Collins, C. & Rollo, E. M. Assessing eating behaviour using upper limb mounted motion sensors: a systematic review. *Nutrients* **11**, 1168 (2019).
30. Doulah, A. et al. The importance of field experiments in testing of sensors for dietary assessment and eating behavior monitoring. in *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5759–5762 (2018).
31. Tricco, A. C. et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann. Intern. Med.* **169**, 467–473 (2018).
32. Veritas Health Innovation. *Covidence Systematic Review Software*, [www.covidence.org](http://www.covidence.org).
33. Bedri, A., Verlekar, A., Thomaz, E., Avva, V. & Starner, T. Detecting mastication: a wearable approach. in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 247–250 (2015).
34. Bedri, A. et al. EarBit: using wearable sensors to detect eating episodes in unconstrained environments. *Proc. ACM Interact., Mob., Wearable Ubiquitous Technol.* **1**, 37 (2017).
35. Bi, S. et al. Auracle: detecting eating episodes with an ear-mounted sensor. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2**, 1–27 (2018).
36. Blechert, J., Liedgruber, M., Lender, A., Reichenberger, J. & Wilhelm, F. H. Unobtrusive electromyography-based eating detection in daily life: a new tool to address underreporting? *Appetite* **118**, 168–173 (2017).
37. Chun, K. S., Bhattacharya, S. & Thomaz, E. Detecting eating episodes by tracking jawbone movements with a non-contact wearable sensor. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2**, 1–21 (2018).
38. Dong, Y., Hoover, A., Scisco, J. & Muth, E. Detecting eating using a wrist mounted device during normal daily activities. in *Proceedings of the International Conference on Embedded Systems, Cyber-physical Systems, and Applications (ESCS)*, 1 (2011).
39. Dong, Y., Scisco, J., Wilson, M., Muth, E. & Hoover, A. Detecting periods of eating during free-living by tracking wrist motion. *IEEE J. Biomed. Health Inform.* **18**, 1253–1260 (2014).
40. Farooq, M., Fontana, J. M., Boateng, A. F., Mccrory, M. A. & Sazonov, E. A comparative study of food intake detection using artificial neural network and support vector machine. *Proc. 2013 12th Int. Conf. Mach. Learn. Appl.* **1**, 153 (2013).
41. Farooq, M. & Sazonov, E. Detection of chewing from piezoelectric film sensor signals using ensemble classifiers. in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 4929–4932 (2016).
42. Farooq, M. & Sazonov, E. Segmentation and characterization of chewing bouts by monitoring temporalis muscle using smart glasses with piezoelectric sensor. *IEEE J. Biomed. Health Inform.* **21**, 1495–1503 (2017).
43. Farooq, M. & Sazonov, E. Accelerometer-based detection of food intake in free-living individuals. *IEEE Sens. J.* **18**, 3752–3758 (2018).
44. Fontana, J. M., Farooq, M. & Sazonov, E. Estimation of feature importance for food intake detection based on Random Forests classification. in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6756–6759 (2013).
45. Fontana, J. M., Farooq, M. & Sazonov, E. Automatic ingestion monitor: a novel wearable device for monitoring of ingestive behavior. *IEEE Trans. Biomed. Eng.* **61**, 1772–1779 (2014).
46. Fortuna, C., Giraud-Carrier, C. & West, J. Hand-to-mouth motion tracking in free-living conditions for improved weight control. in *IEEE International Conference on Healthcare Informatics (ICHI)*, 341–348 (2016).
47. Gao, Y. et al. iHear food: eating detection using commodity bluetooth headsets. in *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 163–172 (2016).
48. Gomes, D. & Sousa, I. Real-time drink trigger detection in free-living conditions using inertial sensors. *Sensors* **19**, 2145 (2019).
49. Hamatani, T., Elhamshary, M., Uchiyama, A. & Higashino, T. FluidMeter: gauging the human daily fluid intake using smartwatches. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2**, 1–25 (2018).
50. Jia, W. et al. Automatic food detection in egocentric images using artificial intelligence technology. *Public Health Nutr.* **22**, 1168–1179 (2019).
51. Kyritsis, K., Diou, C. & Delopoulos, A. Detecting meals in the wild using the inertial data of a typical smartwatch. *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4229–4232 (2019).
52. Mirtchouk, M. et al. Recognizing eating from body-worn sensors: combining free-living and laboratory data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **1**, 1–20 (2017).
53. Navarathna, P., Bequette, B. W. & Cameron, F. Wearable device based activity recognition and prediction for improved feedforward control. in *2018 Annual American Control Conference (ACC)*, 3571–3576 (2018).
54. Rahman, T., Czerwinski, M., Gilad-Bachrach, R. & Johns, P. Predicting “about-to-eat” moments for just-in-time eating intervention. in *Proceedings of the 6th International Conference on Digital Health Conference*, 141–150 (2016).
55. Schiboni, G., Wasner, F. & Amft, O. A privacy-preserving wearable camera setup for dietary event spotting in free-living. in *IEEE International Conference on Pervasive Computing and Communications Workshops*, 872–877 (2018).
56. Sen, S., Subbaraju, V., Misra, A., Balan, R. K. & Lee, Y. Experiences in building a real-world eating recogniser. in *Proceedings of the 4th International on Workshop on Physical Analytics*, 7–12 (2017).
57. Sen, S., Subbaraju, V., Misra, A., Balan, R. & Lee, Y. Annapurna: building a real-world smartwatch-based automated food journal. in *IEEE 19th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 1–6 (2018).
58. Sharma, S., Jasper, P., Muth, E. & Hoover, A. Automatic detection of periods of eating using wrist motion tracking. in *IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 362–363 (2016).
59. Thomaz, E., Essa, I. & Abowd, G. D. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1029–1040 (2015).
60. Thomaz, E., Zhang, C., Essa, I. & Abowd, G. D. Inferring meal eating activities in real world settings from ambient sounds: a feasibility study. in *Proceedings of the 20th International Conference on Intelligent User Interfaces*, 427–431 (2015).
61. Yatani, K. & Truong, K. N. BodyScope: a wearable acoustic sensor for activity recognition. in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 341–350 (2012).
62. Ye, X., Chen, G., Gao, Y., Wang, H. & Cao, Y. Assisting food journaling with automatic eating detection. in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 3255–3262 (2016).
63. Zhang, R. & Amft, O. Free-living eating event spotting using EMG-monitoring eyeglasses. in *IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 128–132 (2018).
64. Zhang, R. & Amft, O. Monitoring chewing and eating in free-living using smart eyeglasses. *IEEE J. Biomed. Health Inform.* **22**, 23–32 (2018).
65. Gemming, L., Utter, J. & Ni Mhurchu, C. Image-assisted dietary assessment: a systematic review of the evidence. *J. Acad. Nutr. Diet.* **115**, 64–77 (2015).
66. Gemming, L., Doherty, A., Utter, J., Shields, E. & Ni Mhurchu, C. The use of a wearable camera to capture and categorise the environmental and social context of self-identified eating episodes. *Appetite* **92**, 118–125 (2015).
67. Vaizman, Y., Ellis, K. & Lanckriet, G. Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE Pervasive Comput.* **16**, 62–74 (2017).
68. Alshurafa, N. et al. Counting bites with bits: expert workshop addressing calorie and macronutrient intake monitoring. *J. Med. Internet Res.* **21**, e14904 (2019).
69. Buchowski, M. S. Doubly labeled water is a validated and verified reference standard in nutrition research. *J. Nutr.* **144**, 573–574 (2014).
70. Hall, K. D. et al. Methodologic issues in doubly labeled water measurements of energy expenditure during very low-carbohydrate diets. *bioRxiv*, <https://doi.org/10.1101/403931> (2018).
71. Alharbi, R. et al. I can't be myself: effects of wearable cameras on the capture of authentic behavior in the wild. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2**, 1–40 (2018).
72. Engel, S. G. et al. Ecological momentary assessment in eating disorder and obesity research: a review of the recent literature. *Curr. Psychiatry Rep.* **18**, 37 (2016).

73. Schembre, S. M. et al. Mobile ecological momentary diet assessment methods for behavioral research: systematic review. *JMIR mHealth uHealth* **6**, e11170 (2018).
74. Spruijt-Metz, D., de la Haye, K., Lach, J. & Stankovic, J. A. M2FED: monitoring and modeling family eating dynamics. in *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems*, 352–353 (2016).
75. Ma, M. et al. M<sup>2</sup>G: a monitor of monitoring systems with ground truth validation features for research-oriented residential applications. in *IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, 10–18 (2017).
76. Leech, R. M., Worsley, A., Timperio, A. & McNaughton, S. A. Characterizing eating patterns: a comparison of eating occasion definitions. *Am. J. Clin. Nutr.* **102**, 1229–1237 (2015).
77. Videon, T. M. & Manning, C. K. Influences on adolescent eating patterns: the importance of family meals. *J. Adolesc. Health* **32**, 365–373 (2003).
78. Neumark-Sztainer, D., Wall, M., Story, M. & Fulkerson, J. A. Are family meal patterns associated with disordered eating behaviors among adolescents? *J. Adolesc. Health* **35**, 350–359 (2004).
79. Suggs, L. S., Della Bella, S., Rangelov, N. & Marques-Vidal, P. Is it better at home with my family? The effects of people and place on children's eating behavior. *Appetite* **121**, 111–118 (2018).
80. Lenardon, J. D., Hansen, A. Y. & Hartley, D. Rural and remote food environments and obesity. *Curr. Obes. Rep.* **4**, 46–53 (2015).
81. Torres, S. J. & Nowson, C. A. Relationship between stress, eating behavior, and obesity. *Nutrition* **23**, 887–894 (2007).
82. Isasi, C. R. et al. Psychosocial stress is associated with obesity and diet quality in Hispanic/Latino adults. *Ann. Epidemiol.* **25**, 84–89 (2015).
83. Sazonov, E. S. & Fontana, J. M. A sensor system for automatic detection of food intake through non-invasive monitoring of chewing. *IEEE Sens. J.* **12**, 1340–1348 (2012).

## ACKNOWLEDGEMENTS

The authors wish to thank Jean Hyon, Anna Satterfield, Rebecca Braganca, and Cynthia Li for their assistance with article review and data extraction. This work was supported in part by National Science Foundation grants SCH-1521740, SCH-1521722, and IIS-1418622; and National Institutes of Health grant 1K25DK113242-01A1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the paper.

## AUTHOR CONTRIBUTIONS

B.M.B. and D.S.-M. contributed to the conception of the review paper. B.M.B. and R.A. contributed to the acquisition of the data. All authors substantially contributed to the analysis and interpretation of data, and to the writing and revision of the paper.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41746-020-0246-2>.

**Correspondence** and requests for materials should be addressed to B.M.B.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020