Resources, Conservation & Recycling 226 (2026) 108689

Contents lists available at ScienceDirect

Resources, Conservation & Recycling

journal homepage: www.sciencedirect.com/journal/resources-conservation-and-recycling

How accurate are food waste tracking systems? Insights from healthcare and hotel kitchens

Leverenz Dominik [a,b,*], Flores Javier [b], Eriksson Mattias [c]

[a] Department of Environmental and Resource Engineering, Waste, Climate & Monitoring, Technical University of Denmark, Kongens Lyngby, 2800, Denmark
[b] Institute for Sanitary Engineering, Water Quality and Solid Waste Management (ISWA), University of Stuttgart, Stuttgart 70569, Germany
[c] Department of Energy and Technology, Swedish University of Agricultural Sciences, Box 7032, Uppsala SE-75007 Uppsala, Sweden

ARTICLE INFO

Keywords:
Food waste tracking
Enumerator bias
Accuracy
Precision
Underreporting
Selfreporting

ABSTRACT

Food waste tracking systems (FWTS) have become increasingly common as monitoring tools in the food service sector. Yet, staff-reported FWTS data are subject to uncertainties from enumerator bias, and their accuracy has rarely been investigated or empirically tested. This study provides insights into the reliability of FWTS data from three healthcare kitchens and one hotel kitchen. Using a staggered experimental design, staff-reported data (≥ 21 days) were compared with scientific control data (6 days) collected with the same FWTS under constant conditions. Staff-reported FWTS data underestimated food waste quantities by up to 80 %, with an average underreporting of approx. 29.4 % across mealtimes (breakfast, lunch, dinner) in the healthcare kitchens and approx. 30.7 % during breakfast buffet in the hotel kitchen. Our results also show that staff-reported FWTS quantities can shift the true mean values toward a biased underestimation without widening the confidence interval, making comparisons appear precise but inaccurate.

1. Introduction

International and regional policy agendas have set ambitious targets for reducing food waste (e.g., United Nations Sustainable Development Goal SDG 12.3 and Waste Framework Directive (Directive 2008/98/EC), which in turn require robust, comparable measurement across all sectors of food value chains (Council of the European Union, 2024; United Nations, 2015). Policies like the EU Circular Economy Action Plan translate this into binding national targets with harmonized monitoring and reporting requirements (European Commission, 2020). For restaurants and food services, recommended methods include direct measurements and waste composition analysis as well as indirect methods like counting and scanning or diaries (European Commission, 2019). In commercial and institutional kitchens, food waste tracking systems (FWTS) have become increasingly popular in recent years (Vardopoulos et al., 2024). Providers with the widest reach and customer base implemented their FWTS in >3000 kitchens (Winnow, 2025) and 4500 kitchens (Leanpath, 2025). These FWTS are designed to collect food waste data at source and gain actionable insights for operational and administrational improvements (Eriksson et al., 2019). The modus operandi ranges from manual logging and weighing to fully automated smart bins and touchless recording with AI-based image recognition and automated data processing. Current developments combine IoT and big-data platforms to connect kitchen sensors and food waste data across sites for benchmarking and reporting (Ahmadzadeh et al., 2023).

1.1. Food waste tracking can drive reductions

Field evidence from the hospitality sector indicates that FWTS can lead to significant food waste reductions due to the operational learnings gained during data collection (Sigala et al., 2025a). For instance, Eriksson et al. (2019) found that among 735 food service providers (hotels, restaurants, and canteens), that systematically tracked their food waste quantities, 61 % subsequently reduced it over time. Leverenz et al. (2021) reported a 64 % decrease in buffet leftovers in four hotels following the installation of food waste-tracking devices. Raised staff awareness led to behavioral adjustments and the independent development of food waste reduction strategies by the kitchen and service staff. Goossens et al. (2022) assessed the economic, environmental, and social benefits of FWTS, providing a business case for food waste tracking tools. Accordingly, food waste was reduced by an average of 1800 kg per kitchen and year, which corresponds to annual net savings

* Corresponding author at: Department of Environmental and Resource Engineering, Waste, Climate & Monitoring, Technical University of Denmark, Kongens Lyngby, 2800, Denmark.
E-mail address: domle@dtu.dk (L. Dominik).

https://doi.org/10.1016/j.resconrec.2025.108689
Received 29 August 2025; Received in revised form 17 October 2025; Accepted 7 November 2025
Available online 13 November 2025

of 8300 EUR and 6.8 tons of CO_2 equivalents (Goossens et al., 2022). Read and Muth (2021) estimate that FWTS can also improve purchasing efficiency and lower costs through better planning, with expected reductions in food purchases of about 3.2–7.6 %. Furthermore, food services might save 7–8 kg of CO_2 equivalents for every dollar invested in waste-tracking (Read and Muth, 2021).

1.2. Data quality and measurement bias

The data quality of FWTS depends not only on device settings, but also on the operational behavior of the kitchen staff that operates the system, which increases the likelihood of measurement biases (Corrado et al., 2019). To our knowledge, however, measurement biases have not been systematically investigated in this context. The quality of self-reported data has only been examined in household settings, where participants consistently underestimate actual food waste quantities (Elimelech et al., 2019b). For instance, Quested et al. (2020) found that household diaries underestimated household food waste between 7 % and 40 % compared to waste compositional analyses. The study identified behavioral reactivity (people wasting less during the diary period), misreporting (not all items discarded being recorded), measurement bias (not all items are weighed), and self-selection bias (those completing a diary being different from the wider population) as key factors contributing to underreporting (Quested et al., 2020). Similarly, van der Werf et al. (2020) found only weak to fair alignment between surveys and measured food waste from a curbside sample of 189 households (van der Werf et al., 2020). Elimelech et al. (2019a) used a hybrid approach that combined a self-assessment questionnaire, a physical waste survey, and a food expenditure survey to triangulate household food waste. The self-assessments hereby slightly underestimated the survey quantities (Elimelech et al., 2019a). However, the findings from household settings indicate that similar reactivity can be expected when kitchen staff operate FWTS in commercial food service settings. Malefors et al. (2019) acknowledge that human factors introduce error in any self-report system. Staff may omit certain waste streams (e.g. liquids down the drain, or waste occurring during peak busy periods) either because it is troublesome to log or to present better figures. Moreover, the frequency and timing of data collection can affect the accuracy and reliability of data collected through self-reporting (Malefors et al., 2019). Infrequent or unscheduled data collection can lead to incomplete or missing data, whereas delayed data reporting can affect the accuracy of collected data, e.g. water losses due to exposure of food waste to the air. Hence, staff-reported FWTS data might lack accuracy and misreport the amount of food waste due to operational shortcomings and enumerator-related bias similar to individuals in households.

1.3. Research gap and contribution of this study

Beyond systems that support food waste tracking, generalizability problems exist due to research design issues and heterogeneity of data collection (Dhir et al., 2020). While cross-sectional studies (e.g. (Sebbane and Costa, 2018)), are common in literature, they are constrained by inherent weaknesses of self-reported surveys, which lack independent verification of accuracy (e.g. (Bharucha, 2018; Liao et al., 2018)). Selection bias could affect representativeness of results, thus leading to false precision of results and increase the risk of misinterpretation (e.g. (Hamerman et al., 2018; McAdams et al., 2019)). Despite the growing number of studies and businesses using FWTS, the data accuracy remains widely unknown. Although studies usually provide confidence intervals for the sample data, they cannot provide information about the actual measurement accuracy of FWTS data due to missing independent control mechanisms or objective audits (Vardopoulos et al., 2024). While a few household studies investigated the gap between self-assessments and measured food waste (e.g. (Elimelech et al., 2019a, 2019b; Quested et al., 2020; van der Werf et al., 2020)), this discrepancy remains unknown for FWTS in the food service sector. In particular, the accuracy of FWTS with respect to enumerator-related bias has not been systematically investigated.

This knowledge gap indicates that existing research may be subject to unknown bias, which could lead to misinterpretation of measurement results, false conclusions, and to misguided reduction strategies. FWTS data may understate actual food waste quantities, and reported reductions may partly reflect changes in recording behavior rather than true decreases. With unknown accuracy, stakeholders and system providers can overstate expected improvements, and cross-site benchmarking may reward underreporting practices rather than real prevention. Where FWTS data feed into policy monitoring and public reporting, unquantified bias can distort baseline performance, overstate year-on-year progress, mis-rank sites or sectors for incentives, and misdirect resources. It can also shape public narratives, for example, statements from providers that "FWTS are proven to halve food waste at scale (Winnow, 2025)" when the effect stems largely from biased measurements or underreporting, creating a false sense of progress.

In this light, our study aims to gain first empirical insights into the accuracy of FWTS data by comparing staff-reported with scientifically controlled measurements in healthcare and hotel kitchens. We aim to estimate the enumerator-related relative bias (relative error) in FWTS data and determine operational correlates of accuracy (e.g., coverage, weighing frequency). By quantifying these errors, we aim to improve practical validation and calibration routines and contribute to more reliable monitoring frameworks that better utilize FWTS for decision-making, reporting, and waste reduction measures in support of sustainability targets.

2. Material and methods

2.1. Study design

Our study uses an experimental design with two staggered measurement periods. In the first period, kitchen staff collected data with the FWTS. This phase is hereafter referred to as staff-reported period. Immediately following the completion of the staff-reported period, researchers collected data a few days later, using the same FWTS under similar day-to-day operational conditions in each kitchen. This phase is hereafter referred to as the scientific control period. The measurements were conducted in three healthcare kitchens of the same company and one hotel kitchen. All participants were introduced to the FWTS protocol through standardized training sessions involving kitchen managers, service staff and chefs. The training sessions covered the operation of the FWTS, the categorization of food waste by mealtime and functional area, and the requirement to avoid operational or administrative changes during the study period to maintain consistent baseline conditions and generate comparable data. Participants were told that the purpose of the study was to contribute to a scientific database on food waste quantities. They were not made aware that the accuracy and reliability of their reporting behavior was the true focus of this study. This intentional withholding of the assessment dimension of the study was intended to minimize behavioral reactivity and allow for a more natural data collection under typical operating conditions. In Phase 1 (staff-reported period), kitchen staff independently logged all food waste data using the FWTS over a period of at least 21 days. This time frame follows the methodology of Eriksson et al. (2019), which recommends a minimum of 21 complete datasets, defined as entries for all main meals or measurement categories on a given day (Eriksson et al., 2019). Due to operational constraints such as rest days, illness, and staffing shortages, staff-reported data were often fragmented, and only complete datasets were retained for comparison. In Phase 2 (scientific control period), following the staff-reported period at each site, trained researchers conducted independent measurements over a six-day period using the same FWTS setup (Fig. 1).

All operational and administrational conditions, including menu

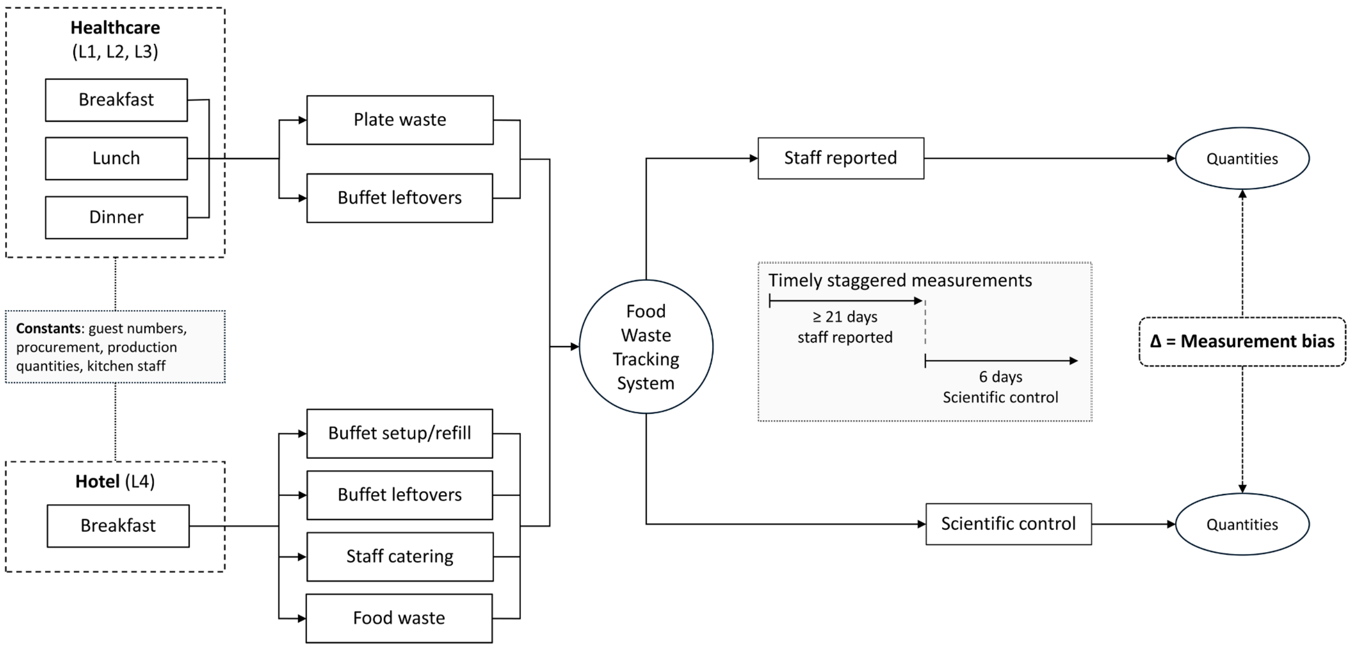

Fig. 1. Study design and process flows.

plans, guest numbers, and procurement routines, were held constant to ensure comparability and to isolate the influence of the data collector (i.e. kitchen staff compared to researchers) as the only systematic difference. The six-day duration represented a practical compromise between methodological rigor and logistical feasibility, since longer scientific audits would have exceeded our staffing and operational capacities. Table 1 provides the characteristics of the measurements, including observation periods, guest occupancy, and completeness of staff-reported versus scientific control FWTS data.

To assess the integrity of the staff-reported data, we used four indicators: (1) temporal gaps, which refers to days or mealtimes without any FWTS entries; (2) the coverage rate, which describes the ratio between the number of items scheduled (i.e., dishes or components listed on the daily menu) and the number of items actually logged (i.e. dishes or components recorded and monitored during the measurement) by the FWTS; (3) the weighing frequency per day (weighings d^{-1}), representing the mean number of FWTS entries per day and meal or operational category; and (4) the relative bias (~enumerator bias), which expresses the proportional difference between the quantities tracked during scientifically controlled measurements and staff-reported measurements (Table 2).

Table 2
Indicators for identifying relative bias in staff-reported food waste data.

No.	Indicator	Definition	Function
(1)	Temporal gaps	Whole days or mealtimes with no data entries.	Indicates time-series integrity
(2)	Coverage rate in % (Items weighed ÷ Items scheduled) ·100′	Proportion of scheduled items [(a)] that were actually recorded.	Indicates skipped dishes or meals when log entries for dishes on menu list are missing.
(3)	Weighing frequency (weighings d^{-1})	Number of FWTS entries per day, mealtime or operational category.	Indicates consistency of reporting
(4)	Relative bias in %	Proportional difference between staff-reported and scientifically controlled measurements.	Quantifies enumerator-related bias (underreporting)

(a). ‘Scheduled’ refers to the total number of food items (dishes or components) listed in the kitchen’s daily menu or buffet plan.

Table 1
Characteristics of the measurements.

Setting	Location (L)	Measurement approach	Occupancy [a] (guests d^{-1}) *	Observation period (dd.mm – dd.mm.yy)	Length [b](days)	Completeness [c] (days)
Care facilities	L1	staff-reported	182 ± 11	15.07 – 01.12.19	138	21
		scientific control	180 ± 7	02.12 – 07.12.19	6	6
	L2	staff-reported	281 ± 8	03.07 – 24.11.19	99	87
		scientific control	280 ± 3	25.11 – 30.11.19	6	6
	L3	staff-reported	501 ± 35	29.07 – 08.09.19	42	31
		scientific control	500 ± 7	09.12 – 14.12.19	6	6
Hotel	L4	staff-reported	114 ± 24	13.03 – 15.07.19	124	27
		scientific control	115 ± 17	16.07 – 28.07.19	12	12

[a] Mean ± standard deviation of registered guests per day. A Mann–Whitney U test showed no significant difference between the staff-reported and scientific-control phases at any site ($p > 0.05$).
[b] Inclusive calendar days between the first and last date shown.
[c] Days on which all mealtimes and waste categories were recorded with a 100 % completeness in the Food-Waste Tracking System (FWTS); i.e. fully documented “waste-measurement days.”.

Given that staff-reported and scientific control measurements took place on different but operationally comparable days, a direct estimate of the absolute measurement bias was not possible. Therefore, we used the relative bias as an indicator, under the assumption that a fairly accurate approximation of absolute bias can be achieved when all operating and administrative conditions are kept constant.

2.2. Food waste tracking procedure and data screenings

All participating kitchens in this study were equipped with a food waste tracking system previously applied in comparable institutional settings (e.g. (Leverenz et al., 2020, 2021). The system uses a calibrated electronic scale with a measurement resolution of 20 g and an intuitive user interface that displays weight values in real-time. The FWTS was integrated into kitchen routines to record food waste promptly and with minimal complexity, comparable to similar FWTS on the market (compare Table S 2). Surplus food was weighed in the same serving trays (bins or dishes) in which the food was originally presented at the buffet. After meal service ended, kitchen staff collected all remaining buffet and plate leftovers and measured them using the FWTS. Pre-configured tare values for each serving tray enabled automated deduction of container weight, ensuring accurate calculation of net food waste. All entries were assigned to specific waste categories (e.g., buffet returns, plate waste) and mealtimes. The data collection protocol hereby complies with the Food Loss and Waste Accounting and Reporting Standard (FLW Standard) established by the World Resources Institute (Hanson et al., 2016). Following a similar approach to Malefors et al. (2019), we characterized each dataset based on quantifiable parameters, such as length of the investigation period, and completeness of data with full resolution or coverage. The quantification period (length) reflects the total number of days during the investigation. Only complete measurement days were used for the comparative data analysis between the staff-reported and the scientific control period. Completeness was reached when FWTS entries were present across all mealtimes during a day, which indicates consistent tracking efforts. Preliminary data screenings revealed entire days without FWTS entries, partial meal coverage, or absence of certain measurement categories (e.g., plate waste vs. buffet returns). Accordingly, temporal gaps and incomplete datasets were excluded from further analysis. This approach corresponds to what Malefors et al. (2019) define as a “second-level” quality criterion, aimed at minimizing analytical noise from partial or inconsistent records. Coverage and weighing frequency were then evaluated for their correlation with relative bias to determine their ability to explain the extent of data inaccuracies.

2.3. Measurement categories

The three healthcare kitchens (L1, L2, L3) belong to the same facility group and follow standardized menu planning and meal preparation practices. Meals in these facilities are served at self-service buffets in dining halls. Food waste in the healthcare facilities was quantified during the three primary mealtimes: breakfast, lunch, and dinner. The hotel, however, does not offer lunch or dinner buffets on a daily basis. Hence, only the breakfast buffet was included in the measurement period in the hotel kitchen. The breakfast buffet follows a largely standardized menu with minor seasonal adjustments, similar to the fixed meal plans in healthcare kitchens. Unlike the healthcare facilities, the hotel kitchen’s food flows were comprehensively quantified, including the food quantities used to set up and refill the buffet, and the buffet leftovers (uneaten and surplus food), which were further distinguished into food waste (the portion of buffet returns that is wasted), and quantities used for other purposes, i.e. staff catering. We expect that operational parameters such as the number of guests had a negligible influence on the FWTS data, as guest occupancy values remained consistent across the staff-reported and scientific control periods (cf. Table 1). Statistical analysis (Mann–Whitney U test) confirmed no significant differences at the 95 % confidence level ($p > 0.05$) for any site.

2.4. Descriptive and inferential statistical analysis

To evaluate the staff-reported food waste data, we compared the daily average food waste quantities (expressed in grams per guest) between the staff-reported and the scientifically controlled measurements. Differences between these paired datasets were analyzed with a two-tailed, paired-sample *t*-test at a 95 % confidence level. Before performing the *t*-tests, the Shapiro-Wilk test revealed normally distributed samples for staff-reported and scientific control measurements after data cleaning. To illustrate differences and distributions within the paired datasets, we provide a set of descriptive plots, including boxplots, scattergrams, strip plots, and bar charts with mean values and standard deviations. For the bias range estimation, we used a resampling procedure with replacement (non-parametric bootstrap method). With this approach we standardized sample sizes across groups to plot the full range of possible differences between staff-reported and scientifically controlled data. We generated 1000 resampled values with replacement for each dataset (staff-reported and controlled) to form synthetic samples of equal size. We then computed the pairwise differences between all possible combinations (yielding 1000×1000 difference values). These differences were then used to construct empirical distributions and calculate 95 % confidence intervals. The bootstrap analysis was applied separately to each mealtime (breakfast, lunch, dinner) in the healthcare kitchens, as well as to aggregated daily values. In the hotel kitchen, the analysis was stratified by four operational sub-categories: buffet setup and refill (1), buffet returns (2), staff catering (3), and food waste (4).

3. Results

3.1. Accuracy of FWTS and underreporting patterns

The results of our study reveal a consistent pattern of underreporting in the staff-reported food waste data collected from the pilot kitchen settings. Across all kitchens in healthcare and hotel environments, the quantities of food waste documented by kitchen staff fell below those obtained via scientifically controlled measurements, with no overreporting in any setting. Fig. 2 categorizes food waste quantities by mealtime (breakfast, lunch, dinner) and provides a daily average for the three healthcare kitchens (L1, L2, L3). The staff-reported accuracy (light blue) is shown as a proportion of the scientifically controlled values. A value of 100 % indicates perfect alignment between staff-reported and scientifically controlled data, whereas values below 100 % indicate underreporting (dark blue). Only one kitchen (L2) showed accurate and reproducible data across all mealtimes. For the other healthcare kitchens (L1 and L3), the daily averages across all mealtimes were underestimated by between 17 % (L3) and 70 % (L1).

In the hotel kitchen, substantial underreporting was observed across all measurement categories (Fig. 3 and Table S 4). The level of underreporting within different measurement categories appeared to align with operational priorities and perceived product values. Reporting was most accurate for buffet setup and refilling, with an underreporting of approximately 10 %. Categories that typically have less operational priority demonstrated more substantial inaccuracies. For instance, staff catering showed an average underreporting of approx. 30 %, buffet returns approx. 33 %, and food waste approx. 49 %. In both the hotel

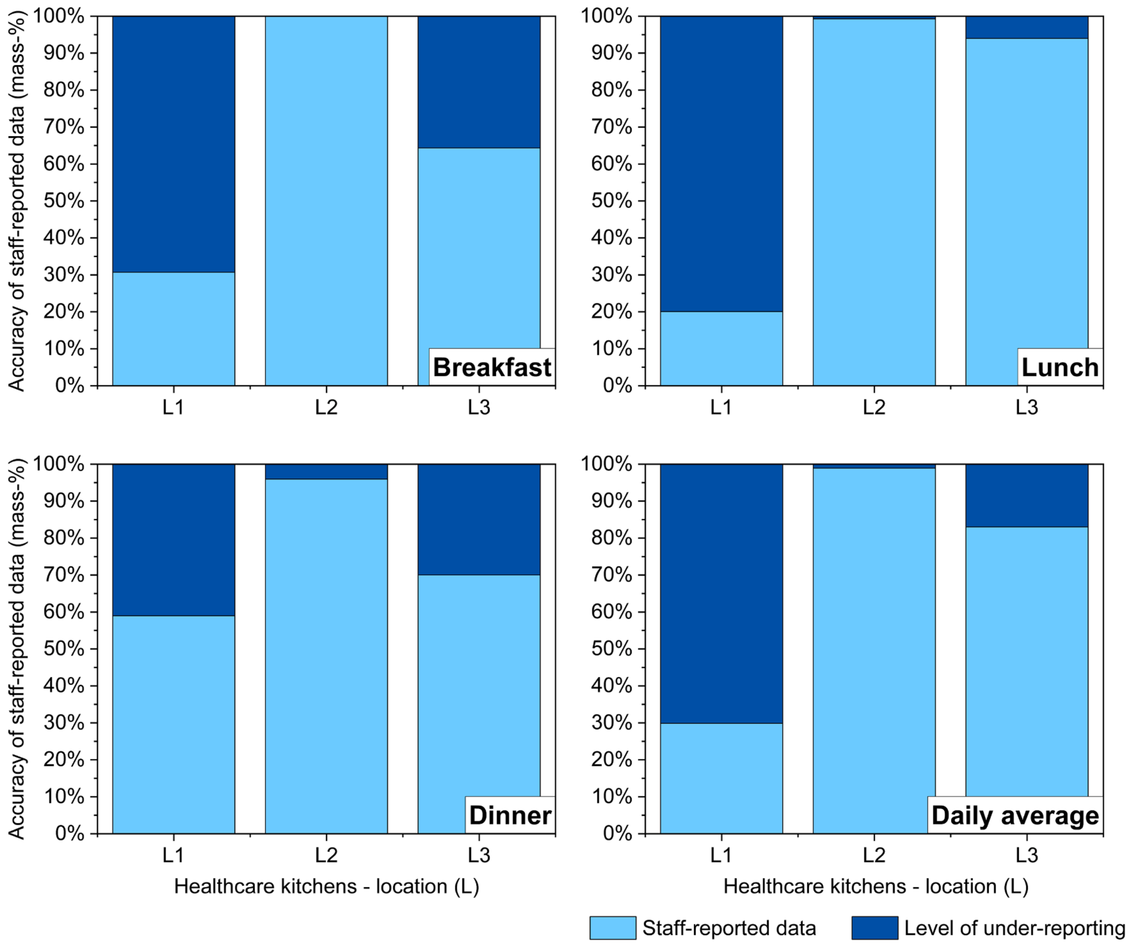

Fig. 2. Levels of underreporting by mealtime in the healthcare kitchens (L1, L2, L3).

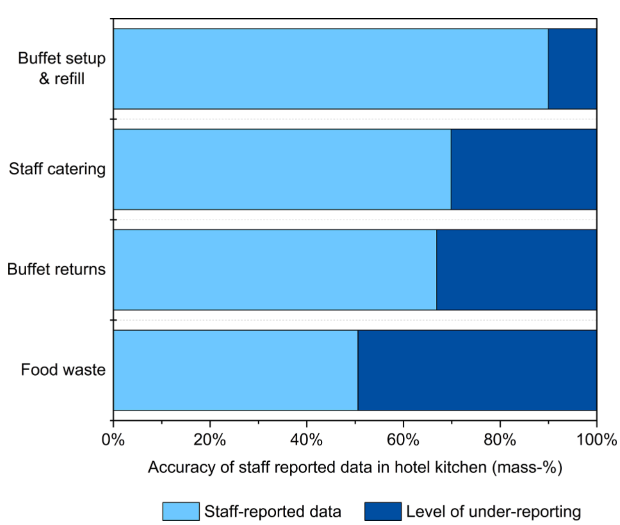

Fig. 3. Underreporting levels of food flows at breakfast buffet in the hotel kitchen (L4).

and healthcare kitchens, statistically significant differences ($p < 0.05$) were found between staff-reported and scientifically controlled measurements for all locations and all mealtimes, except at location L2 during breakfast (Figure S 1).

3.2. *Coverage and weighing frequency as indicators for the magnitude of underreporting*

Relative bias was strongly correlated with both weighing frequency and coverage (cf. Fig. 4). Higher frequencies and broader coverage reduced bias and brought reported values closer to the scientific controls, whereas low values of either were associated with substantially higher bias. For healthcare kitchens, polynomial regressions explained 80 % of the variance in relative bias for weighing frequency (Fig. 4, a) and 82 % for coverage (Fig. 4, b). In the hotel kitchen, the in-sample fit was even stronger, with $R^2 = 0.90$ (Fig. 4, c) and $R^2 = 0.95$ (Fig. 4, d). This stronger statistical correlation is probably not only the result of consistent operating conditions, but also of enumerator effects that can be attributed to the same group of employees in the hotel during a single meal.

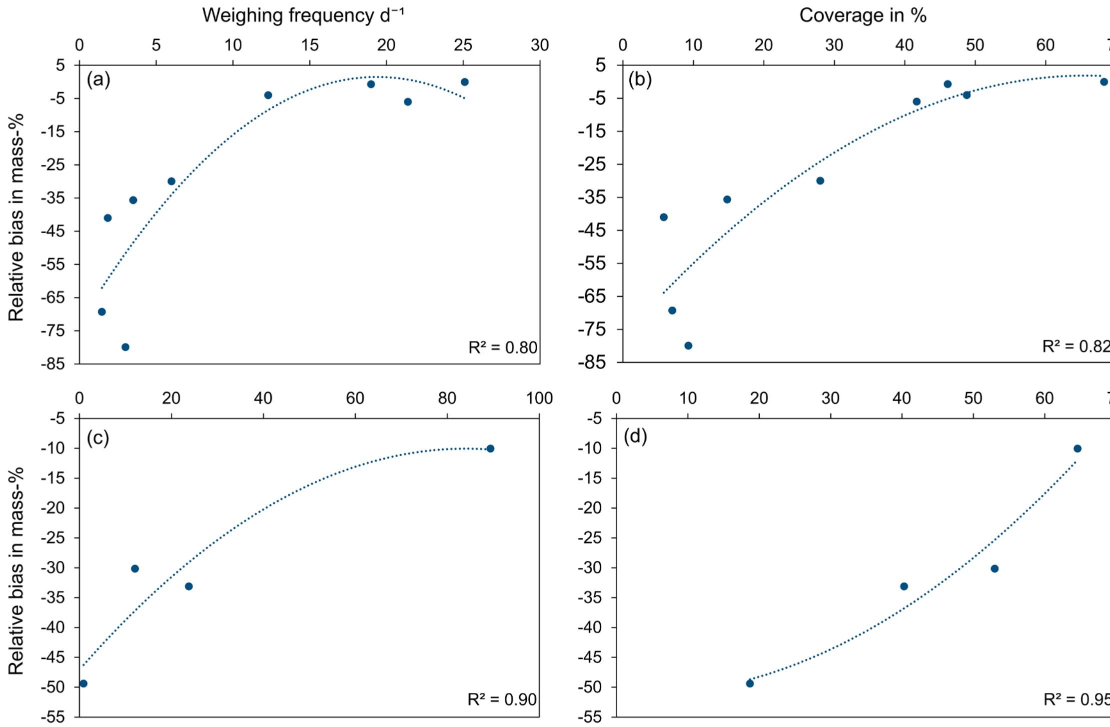

Fig. 4. Polynomial regressions relating relative bias (underreporting) to weighing frequency in healthcare (a) and hotel (c) kitchens, and to coverage in healthcare (b) and hotel (d) kitchens.

3.3. Distributions of relative bias: assessing the range and variability of underreporting

The range of the relative bias, expressed as underreporting distributions, varied across kitchens, mealtimes, and operational categories (cf. Figs. 5 and 6). The density distribution curves demonstrate that narrow confidence intervals do not necessarily imply accurate data. For instance, breakfast and lunch at L1 show the highest underreporting between approx. 60 % and 90 % with relatively narrow intervals. Such narrow intervals make the data appear statistically accurate in the original FWTS data, but do not reveal that the estimates are systematically biased compared the true mean.

Dinner at L1 showed intervals with greater variability between 0 % to 70 %, but with lower relative bias. In contrast, there was no significant underreporting in L2, but the confidence intervals were the widest compared to the other health kitchens. This shows that wider confidence intervals do not necessarily indicate higher degree of underreporting or greater relative bias, but possibly the opposite. Staff reports can produce narrow intervals around systematically underestimated means, making data look statistically reliable while in fact being substantially inaccurate.

In the hotel kitchen, food waste was the most underreported category and had the widest confidence interval, ranging from 20 % to 80 % (Fig. 6). The buffet setup and refilling category showed the lowest underreporting (approx. 10 %) with a correspondingly wide range between –20 % and +35 %. Buffet returns and staff catering categories fell in between, with distributions between around 10 % and 50 % of underreporting. These findings from the hotel kitchen further strengthen the observation that confidence intervals do not indicate data quality or accuracy of FWTS.

4. Discussion

4.1. Main findings: underreporting patterns and data quality

This study is one of the first that systematically investigates the data accuracy of FWTS in the food service sector. We observed consistent and substantial underreporting of staff-reported FWTS data in hotel and healthcare kitchens, caused by an enumerator bias. On average, healthcare kitchens showed an underreporting rate of approximately 29.4 % across all locations and all mealtimes, while hotel kitchens showed an underreporting rate of 30.7 % across all operational categories during breakfast. These underreporting levels are similar compared to self-reporting studies in households (e.g. (Hoover and Moreno, 2017; McDermott et al., 2018; WRAP, 2009, 2013)), where food waste diaries underestimated the true quantities between 7 % and 40 % of mass (Quested et al., 2020). The average underreporting across these five household studies was approx. 30 %, which is nearly identical to the average underreporting in healthcare and hotel kitchens from our study. Furthermore, Elimelech et al. (2019b) observed that subjective self-assessments in households (diaries and questionnaires) showed slightly higher underestimations (approx. 16 %) compared to physical food waste surveys. These findings indicate that the consistent pattern of underreporting persists in different contexts, whether in individual households or in larger operations, i.e. commercial or institutional kitchens.

In our study, however, the level of underreporting differed across locations (kitchens), mealtimes and operational categories. Except for one healthcare kitchen (L2), staff-reported quantities were significantly lower than those compared to scientifically controlled measurements across healthcare and hotel settings. In three out of four pilot kitchens (L1, L3, L4), staff-reported FTWS data significantly underestimated true quantities. The trends of underreporting were statistically significant ($\alpha = 0.05$) across both settings (healthcare and hotel kitchens), all mealtimes (breakfast, lunch, and dinner), and all measurement categories (buffet setup & refill, food waste, buffet returns and staff catering). In hotels, the category of buffet setup and refilling showed relatively low bias (approx. 10 %), while underreporting for buffet returns (33 %) and food waste (49 %) was significantly higher. The latter often involves more variable and less predictable elements, making accurate reporting more challenging and susceptible to underreporting. Hence, the operational context and product values influence the reliability of staff-

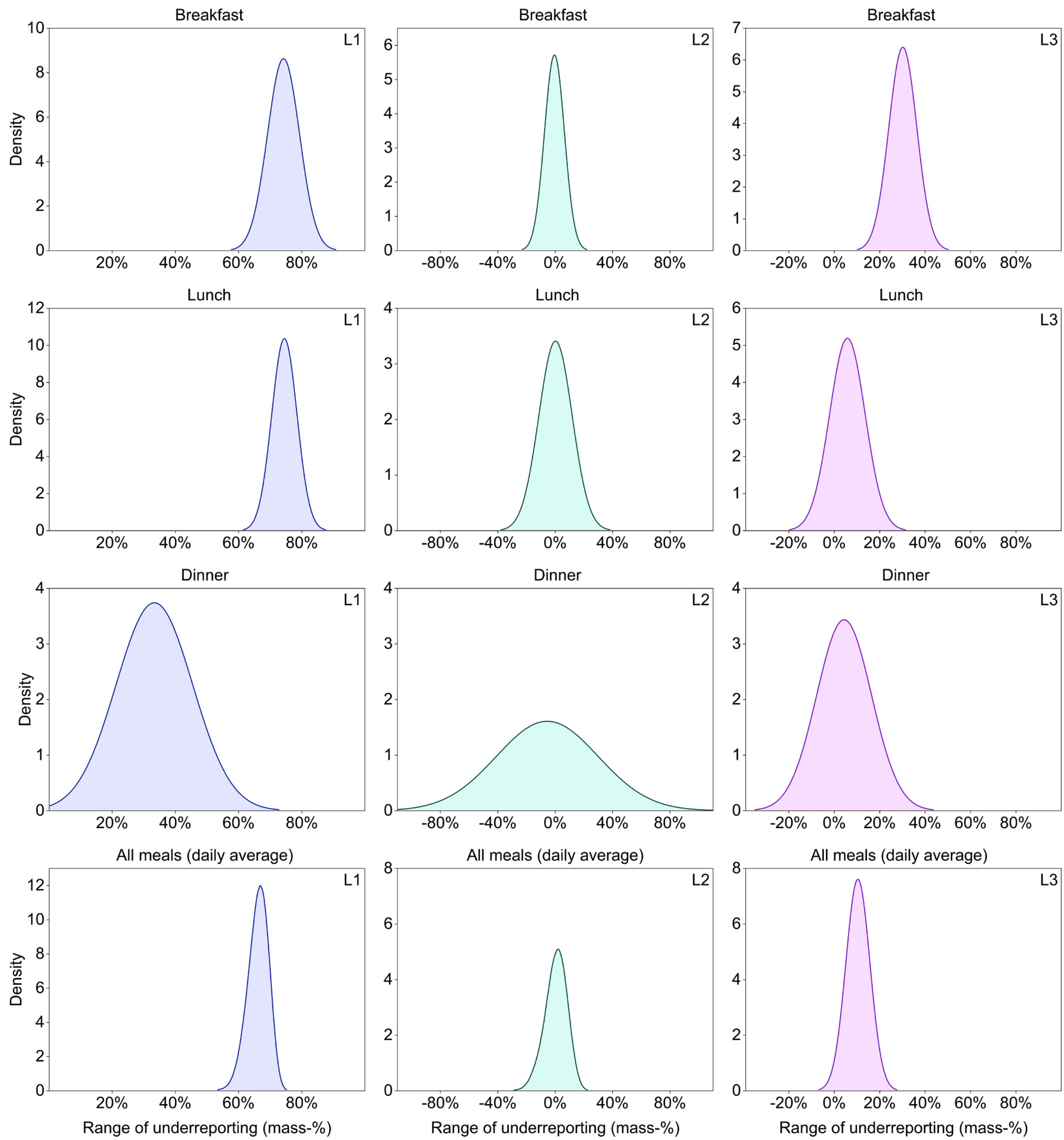

Fig. 5. Range of underreporting (% of mass) by mealtime for healthcare kitchens; columns: L1, L2, L3; rows: breakfast, lunch, dinner, daily average.

reported data. In environments where shifts are short-staffed or responsibilities are poorly defined, measurement failures and inconsistencies become more likely due to the increasing error propagation probability. Similar to Eriksson et al. (2019), we found incomplete entries, missing categories or gaps for whole days in the FTWS logs. We observed that data accuracy correlates more strongly with the location of the kitchen than with its type (e.g., healthcare or hotel kitchen). For instance, even within healthcare kitchens belonging to the same facility group and using standardized meal planning and prepping schemes, the level of underreporting differed significantly between locations. The daily averages, for example, underestimated food waste quantities by ∼ 1 % at Location 2 (L2), ∼ 17 % at Location 3 (L3), and ∼ 70 % at Location 1 (L1). These discrepancies indicate the impact of enumerator-related factors at each location, such as personnel capacities, receptiveness to procedural instructions (training), and the motivation of staff responsible for food waste tracking. The extent of relative bias varied also by mealtime within the same site, which can be explained by shift-specific staff changes. Consequently, our study shows that the individual behavior and practices of employees at each location appear to be one of the main factors influencing the data quality of FWTS.

The highest underreporting was observed in healthcare kitchens during lunch in location L1, with relative biases reaching approx. 80 %. In contrast, location L2 displayed minimal discrepancies between staff-reported and controlled measurements, demonstrating that FWTS can produce reliable data without any notable bias. Importantly, the

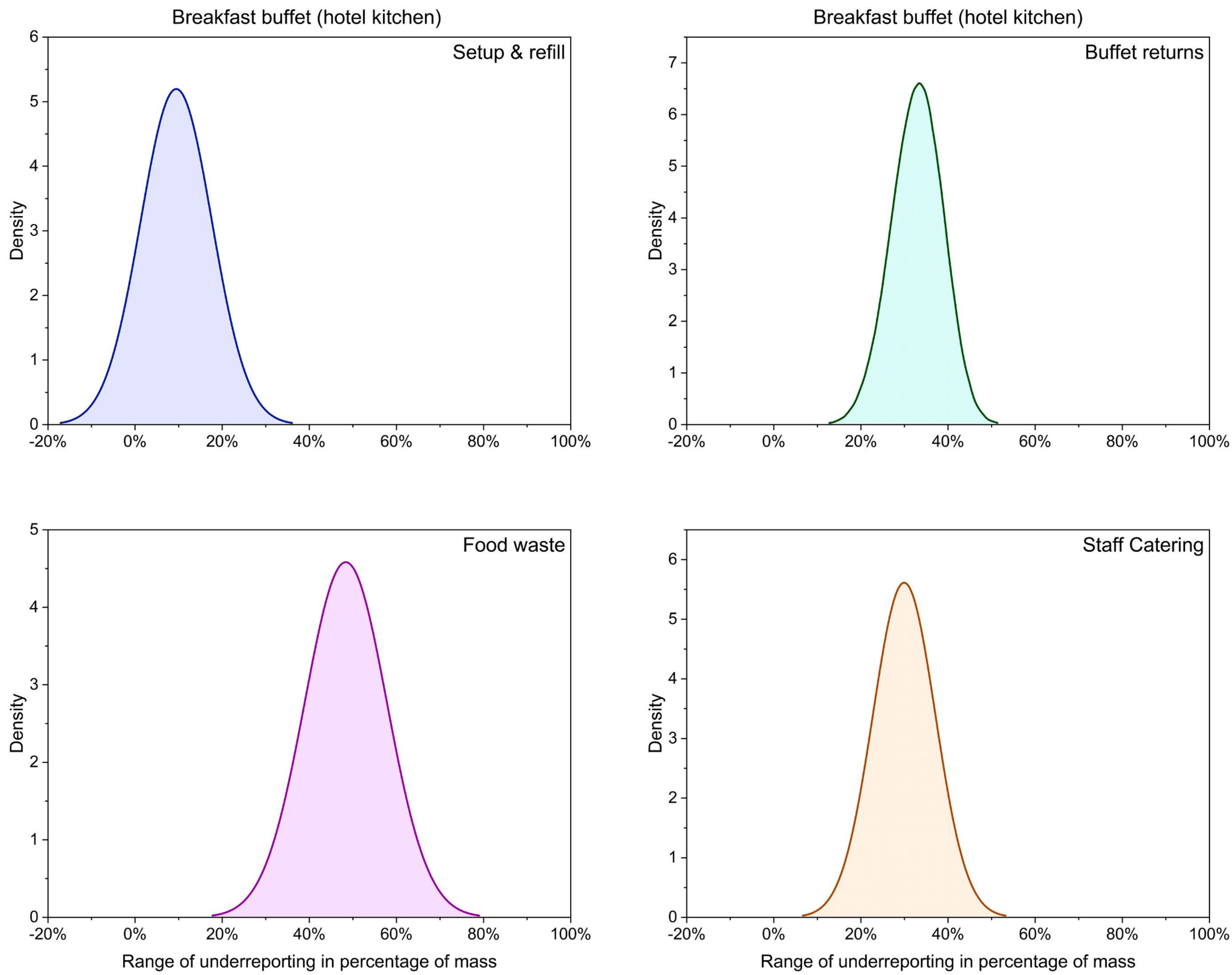

Fig. 6. Range of underreporting (% of mass) by measurement category in the hotel kitchen; top left: buffet setup & refill; top right: buffet returns; bottom left: food waste; bottom right: staff catering.

presence of measurement error does not compromise the usefulness of FWTS. As shown by Eriksson (2019) and Leverenz et al. (2021), FWTS often lead to meaningful waste reductions through awareness-raising, even when absolute quantities remain uncertain. Therefore, while underreporting undermines data accuracy, it may simultaneously reflect a positive behavioral shift triggered by measurement itself, as exemplified in Sundin et al. (2024). Hence, FWTS can serve not only as monitoring tools but also as behavioral interventions, provided that their limitations are recognized and mitigated through appropriate validation strategies. To identify underreporting in daily kitchen processes, operational parameters like weighing frequency and coverage can serve as practical indicators due to their correlation to the relative bias. In healthcare settings, polynomial regression (in-sample coefficient of determination, R^2) explained 80 % of the variance in the relative bias for weighing frequency and 82 % for coverage. In the hotel setting, the corresponding R^2 values were 0.90 and 0.95, indicating even stronger in-sample fit. These results suggest substantial associations or correlations in the observed data. However, high in-sample R^2 does not equal generalizability. The small sample size and potential single-location effects increase the risk of overfitting and confounding (including multicollinearity among polynomial terms). Accordingly, we recommend using coverage and weighing frequency as practical indicators in real-world settings, serving more as a rule of thumb rather than a direct calibrator for the level of underreporting.

4.2. Theoretical and practical implications

The observed underreporting tendency in this study reflects a complex interaction between task design, individual motivation, and institutional context. While FWTS may be framed as important tools to monitor and reduce food waste (e.g. (Eriksson et al., 2019; Goossens et al., 2022; Malefors et al., 2024; Orr and Goossens, 2024; Principato et al., 2023; Sigala et al., 2025)), their implementation in high-pressure kitchen environments introduces a subjective dimension that can compromise data quality. The results of our study indicate that FWTS data in the current literature may understate actual food waste quantities (e.g. (Diana et al., 2023; Filimonau and De Coteau, 2019; Vardopoulos et al., 2024)) and that the reported reductions may partly reflect changes in recording behavior rather than actual declines (e.g. (Eriksson et al., 2019; Goossens et al., 2022; Leverenz et al., 2021; Sigala et al., 2025a)). Without objective audits, reported improvements might be significantly overstated and cross-site benchmarking may reward tracking practices rather than real prevention. Where FWTS data feed into policy monitoring and reporting, unquantified bias can overstate year-on-year progress and shape public narratives. Advertising statements of kitchen providers that “FWTS are proven to halve food waste

(Winnows 2025)" need to be treated with caution. Touchless systems such as smart bins might be less vulnerable to enumerator bias and help to reduce the level of underreporting because they depend more on the accuracy of their components, such as the precision of the scale or image recognition modules (Malefors et al., 2024). Although automation is helpful, it is not a substitute for human engagement. Our findings showed that technology alone might not resolve reporting errors if the personnel interacting with the systems lack the capacity, time, clarity or incentive to report accurately. In this light, we suggest an integrative approach to improve data quality of FWTS by combining control mechanisms regarding measurement consistency (e.g. using coverage and weighing frequency as rule-of-thumb-indicators) with frequent staff training to ensure procedural clarity. FWTS implementations should consider enumerator-related biases and their impact on system effectiveness. When supported by diagnostic indicators and periodic validation, FWTS can provide reliable data for food waste monitoring and reduction strategies, as demonstrated by the accurate data from one healthcare kitchen (L2) in our study. However, considering the predominant pattern of significant underreporting, the use for external reporting needs to be treated with caution.

4.3. Understanding the sources of enumerator bias in staff-reported data

The observed underreporting pattern in this study can be attributed to various interrelated behavioral, operational, and methodological factors. Several studies have shown that the act of measurement itself can influence the outcome and lead to food waste reductions, but without considering potential enumerator biases (den Boer et al., 2024; Elimelech et al., 2019a, 2019b; Eriksson et al., 2019; Goossens et al., 2022; Leverenz et al., 2019, 2020, 2021; Quested et al., 2020). Broderick & Gibson (2019), however, acknowledge that the Hawthorne effect likely influenced the kitchen staff behavior during their field audits toward underestimation. Accordingly, the observer effect may have led kitchen staff to systematically underreport the amount of food waste in order to portray themselves or their team as more efficient or environmentally conscious. This might also be related to social desirability, which means that people tend to adjust their stated behavior to meet perceived social expectations or personal ideals (van Herpen et al., 2019). For example, in a workplace cafeteria, large discrepancies were found between stated amounts within a survey and actual behaviors, i.e. observed amounts of leftovers (Sebbane and Costa, 2018). Discrepancies were predicted by descriptive norms and were more common among men (1.5 times more likely than women) despite smaller waste amounts, indicating that social expectations shape self-reports as much as behavioral aspects. Procedural clarity might further affect data collection, especially when different staff members conduct reporting procedures inconsistently, like in the healthcare kitchens in our study. For FWTS, this relates to technical aspects such as accuracy and calibration of the scale, but also to operational procedures such as correct categorization of the food and waste items, completeness of measurements, or even double entries. Furthermore, operational pressures during peak service periods might lead to missed or incomplete entries (like in our study), when kitchen staff prioritizes service continuity over FWTS data accuracy. While Malefors et al. (2019) provide a useful day-count rule for attaining a measurement precision of the "Waste per portion", their calculation assumes unbiased, independent measurement errors (Malefors et al., 2019). Our results, however, showed that in staff-reported FWTS data, enumerator bias (e.g., selective recording, inconsistent inclusion rules) can shift the true value toward substantial underestimation without widening the confidence interval. That means that longer measurement periods and larger samples may narrow intervals around biased estimates, making inaccurate results appear precise. Therefore, staff-reported FWTS data can produce narrow intervals around systematically underestimated means, making data look statistically reliable while in fact being substantially inaccurate. Consequently, cross-site or temporal comparisons may appear precise but should be interpreted with caution. Where feasible, independent audits or models that minimize enumerator effects should be used to align reported precision with true measurement uncertainty.

4.4. Limitations

The design of our study did not allow for simultaneous dual-operator measurement (i.e., staff and researcher weighing concurrently) due to practical constraints in personnel and logistics. This decision was also informed by concerns that researcher presence could introduce reactivity and influence normal operations. As a compromise, a temporally staggered design was implemented, with scientifically controlled measurements conducted shortly after the staff-reported phase under stable conditions. While this design does not allow to quantify the absolute bias, it offers a pragmatic and minimally invasive approach to isolating the effect of the reporting enumerator and calculate the relative bias. Throughout the study period, critical operational parameters, such as guest numbers, menu plans, and procurement procedures, remained consistent, increasing the reliability of comparisons. Hence, first inferences were drawn about the existence, direction and relative extent of systematic enumerator-related bias. This approach represents a pragmatic response to the logistical constraints of institutional kitchen environments and research capacities, while offering a first indication of the extent to which enumerator effects may compromise the reliability of staff-reported food waste data. However, sample size and regional specificity of the participating kitchens might restrict the generalizability of the results beyond the context studied. The focus on healthcare and hotel kitchens also limits the applicability of findings to other sectors, such as schools, catering services, or institutional canteens, where food service practices and reporting structures may differ. External factors such as seasonality, guest composition, or unforeseen operational changes (e.g., staff absences, menu shifts) could have influenced waste quantities, although efforts were made to stabilize these parameters across the measurement periods. Despite these limitations, the observed levels of underreporting align with findings from household studies (Elimelech et al., 2019b; Quested et al., 2020), suggesting a cross-contextual validity of the bias. To our knowledge, this study is among the first to quantify underreporting in institutional kitchens using controlled comparisons, offering an important empirical foundation for future research.

4.5. Recommendations for future research

Given the inherent challenges in aligning staff-reported data with scientific rigor, future investigations might contribute to investigating possibilities in reducing the enumerator-related bias. Additional attention should be given to contextual variables such as meal types, service formats, kitchen layout, and staffing levels. As observed in this study, even within standardized operational frameworks, site-specific variations in underreporting were substantial. Research aimed at characterizing and modeling these contextual factors could help develop predictive tools for data quality risk assessment. To enhance methodological rigor, hybrid designs could combine staff reports with periodic scientific validation during dual measurement phases or randomized spot audits. The integration of sensor-based automation tools, e.g. image recognition scales or IoT-enabled waste bins, offers technological possibilities for reducing human error and increasing data reliability. Some of the FWTS on the market already provide those features (cf. Table S 2). Interventional studies could assess how procedural improvements and digital feedback loops influence both reporting behavior and waste reduction outcomes over time. Refining FWTS protocols to incorporate diagnostic metrics and real-time validation could contribute to improving both the operational and scientific value of food waste tracking in institutional settings.

5. Conclusion

This study offers one of the first systematic, controlled assessments of the data accuracy of food waste tracking systems in hotel and healthcare kitchens. Across sites, staff-reported data consistently underestimated true quantities by approx. 30 %. Despite standardized menus and operational procedures, reporting accuracy varied widely across sites, mealtimes and measurement categories, indicating that operating personnel have the greatest influence on data quality. Notably, one healthcare kitchen (L2) produced near-unbiased data, demonstrating that FWTS can yield reliable measurements when well implemented. We identified operational drivers of data quality, showing strong in-sample associations between weighing frequency/coverage with relative bias (healthcare: $R^2 \approx 0.80/0.82$; hotel: $R^2 \approx 0.90/0.95$). Scientifically, the results shift attention from random errors to enumerator bias, showing that systematic underreporting is a primary threat to data validity of FWTS. Narrow confidence intervals can coexist with biased means, making inaccurate data seem robust. The work separates behavioral (enumerator) from technical error sources and offers field-tested indicators to diagnose risk. For kitchen managers and FWTS providers, data quality improves when implementations combine procedural clarity and role ownership with simple diagnostic metrics (track and act on weighing frequency and coverage thresholds). For benchmarking, corporate reporting, and policy monitoring, our findings raise concerns for unvalidated cross-site comparisons. Periodic independent audits or randomized checks should be built into FWTS procedures. For technology adopters, automation (e.g., smart bins, touchless systems) might reduce enumerator bias but still require training, oversight, and validation to ensure accurate food waste tracking.

Funding

This work was supported by the German Federal Ministry of Food and Agriculture, funding the ELOFOS project [grant number: 281A103516].

CRediT authorship contribution statement

Leverenz Dominik: Writing – original draft, Formal analysis, Data curation. **Flores Javier:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Eriksson Mattias:** Writing – review & editing, Writing – original draft, Validation, Data curation.

Declaration of competing interest

Dominik Leverenz contributed to the development of the food waste tracking system (Resourcemanager Food), which was used for data collection in this study. The application is freely available in the Google Play store, has no commercial purpose, and the copyright is held by the University of Stuttgart. Mattias Eriksson was one of the founders and part-owner of Matomatic AB, a spinoff from the Swedish University of Agricultural Sciences. No data or financial support from this company were involved in the present study. The authors declare no other financial or personal relationships that could have influenced the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the cooperation of the project partners from the healthcare and hotel kitchens, whose participation and provision of information made this research possible. We also thank our student research assistants, Pornpimon Siripanaroj, Shuk Man Yu, and Simon Rauh, for their important support during the measurements and data analysis.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.resconrec.2025.108689.

Data availability

Data will be made available on request.

References

Ahmadzadeh, S., Ajmal, T., Ramanathan, R., Duan, Y., 2023. A comprehensive review on food waste reduction based on IoT and big data technologies. Sustainability 15 (4), 3482. https://doi.org/10.3390/su15043482.

Bharucha, J., 2018. Tackling the challenges of reducing and managing food waste in Mumbai restaurants. Br. Food J. 120 (3), 639–649. https://doi.org/10.1108/BFJ-06-2017-0324.

Broderick, S., Gibson, C., 2019. Reducing commercial food waste in Ireland. https://www.epa.ie/publications/research/waste/Research_Report_282.pdf.

Corrado, S., Caldeira, C., Eriksson, M., Hanssen, O.J., Hauser, H.-E., van Holsteijn, F., et al., 2019. Food waste accounting methodologies: challenges, opportunities, and further advancements. Glob. Food Sec. 20, 93–100. https://doi.org/10.1016/j.gfs.2019.01.002.

Council of the European Union. (2024). *Proposal for a Directive of The European Parliament and of the Council amending Directive 2008/98/EC on waste*. https://data.consilium.europa.eu/doc/document/ST-11300-2024-INIT/en/pdf.

den Boer, J., Skiba, A., den Boer, E., Obersteiner, G., Dyjakon, A., 2024. Prevention and under-reporting effects of food waste diaries. Sustainability 16 (24), 11009. https://doi.org/10.3390/su162411009.

Dhir, A., Talwar, S., Kaur, P., Malibari, A., 2020. Food waste in hospitality and food services: a systematic literature review and framework development approach. J. Clean. Prod. 270, 122861. https://doi.org/10.1016/j.jclepro.2020.122861.

Diana, R., Martianto, D., Baliwati, Y.F., Sukandar, D., Hendriadi, A., 2023. Food waste in Indonesian hospitals: a systematic review. Nutr. Food Sci. 53 (5), 881–900. https://doi.org/10.1108/NFS-05-2022-0150.

Elimelech, E., Ert, E., Ayalon, O., 2019a. Bridging the gap between self-assessments and measured household food waste: a hybrid valuation approach. Waste Manag. 95, 259–270. https://doi.org/10.1016/j.wasman.2019.06.015.

Elimelech, E., Ert, E., Ayalon, O., 2019b. Exploring the drivers behind self-reported and measured food wastage. Sustainability 11 (20), 5677. https://doi.org/10.3390/su11205677.

Eriksson, M., Malefors, C., Callewaert, P., Hartikainen, H., Pietilinen, O., Strid, I., 2019. What gets measured gets managed or does it? Connection between food waste quantification and food waste reduction in the hospitality sector. Resour. Conserv. Recycl.: X 4, 100021. https://doi.org/10.1016/j.rcrx.2019.100021.

European Commission, 2019. Pinimum quality requirements for the uniform measurement of levels of food waste. http://data.europa.eu/eli/dec_del/2019/1597/oj.

European Commission, 2020. A new circular economy action plan: for a cleaner and more competitive Europe (COM(2020) 98 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0098.

Filimonau, V., De Coteau, D.A., 2019. Food waste management in hospitality operations: a critical review. Tour. Manag. 71, 234–245. https://doi.org/10.1016/j.tourman.2018.10.009.

Goossens, Y., Leverenz, D., Kuntscher, M., 2022. Waste-tracking tools: a business case for more sustainable and resource efficient food services. Resour. Conserv. Recycl. Adv. 15, 200112. https://doi.org/10.1016/j.rcradv.2022.200112.

Hamerman, E.J., Rudell, F., Martins, C.M., 2018. Factors that predict taking restaurant leftovers: strategies for reducing food waste. J. Consum. Behav. 17 (1), 94–104. https://doi.org/10.1002/cb.1700.

Hanson, C., Lipinski, B., Robertson, K., Dias, D., Gavilan, I., Gréverath, P., Ritter, S., Fonseca, J., van Otterdijk, R., Timmermans, T., Lomax, J., O'Connor, C., Dawe, A., Swannell, R., Berger, V., Reddy, M., Somogyi, D., Tran, B., Leach, B., Quested, T., 2016. Food loss and Waste Accounting and Reporting Standard. https://flwprotocol.org/wp-content/uploads/2017/05/FLW_Standard_final_2016.pdf.

Hoover, D., Moreno, L., 2017. Estimating quantities and types of food waste at the city level. https://www.nrdc.org/sites/default/files/food-waste-city-level-report.pdf.

Leverenz, D., Hafner, G., Moussawel, S., Kranert, M., 2021. Reducing food waste in hotel kitchens based on self-reported data. Ind. Mark. Manag. 93, 617–627. https://doi.org/10.1016/j.indmarman.2020.08.008.

Leanpath, 2025. Slash your food waste and boost profits now. https://www.leanpath.com/.

Leverenz, D., Moussawel, S., Hafner, G., Kranert, M., 2020. What influences buffet leftovers at event caterings? A German case study. Waste Manage 116, 100–111. https://doi.org/10.1016/j.wasman.2020.07.029.

Leverenz, D., Moussawel, S., Maurer, C., Hafner, G., Schneider, F., Schmidt, T., Kranert, M., 2019. Quantifying the prevention potential of avoidable food waste in households using a self-reporting approach. Resour. Conserv. Recycl. 150, 104417. https://doi.org/10.1016/j.resconrec.2019.104417.

Liao, C., Hong, J., Zhao, D., Zhang, S., Chen, C., 2018. Confucian culture as determinants of consumers' Food leftover generation: evidence from Chengdu, China. Environ. Sci. Pollut. Res. 25 (15), 14919–14933. https://doi.org/10.1007/s11356-018-1639-5.

Malefors, C., Callewaert, P., Hansson, P.-A., Hartikainen, H., Pietiläinen, O., Strid, I., Strotmann, C., Eriksson, M., 2019. Towards a baseline for food-waste quantification in the hospitality sector—Quantities and data processing criteria. Sustainability 11 (13), 3541. https://doi.org/10.3390/su11133541.

Malefors, C., Svensson, E., Eriksson, M., 2024. Automated quantification tool to monitor plate waste in school canteens. *Resources, Conservation & Recycling, 200*, 107288. https://doi.org/10.1016/j.resconrec.2023.107288.

McAdams, B., von Massow, M., Gallant, M., Hayhoe, M.-A., 2019. A cross industry evaluation of food waste in restaurants. J. Foodserv. Bus. Res. 22 (5), 449–466. https://doi.org/10.1080/15378020.2019.1637220.

McDermott, C., Elliott, D., Johnson, A., Hunter, K., & de Venecia, C. (2018). *2017 Oregon wasted food Study: residential sector waste sort, diary, and survey study*. https://www.oregon.gov/deq/mm/Documents/ResKitchenDiarySurvey.pdf.

Orr, L., Goossens, Y., 2024. Trimming the plate: a comprehensive case study on effective food waste reduction strategies in corporate canteens. Sustainability 16 (2), 785. https://doi.org/10.3390/su16020785.

Principato, L., Marchetti, S., Barbanera, M., Ruini, L., Capoccia, L., Comis, C., Secondi, L., 2023. Introducing digital tools for sustainable food supply management: tackling food loss and waste in industrial canteens. J. Ind. Ecol. 27 (4), 1060–1075. https://doi.org/10.1111/jiec.13391.

Quested, T.E., Palmer, G., Moreno, L.C., McDermott, C., Schumacher, K., 2020. Comparing diaries and waste compositional analysis for measuring food waste in the home. J. Clean. Prod. 262, 121263. https://doi.org/10.1016/j.jclepro.2020.121263.

Read, Q.D., Muth, M.K., 2021. Cost-effectiveness of four food waste interventions: is food waste reduction a "win–win?". Resour. Conserv. Recycl. 168, 105448. https://doi.org/10.1016/j.resconrec.2021.105448.

Sebbane, M., Costa, S., 2018. Food leftovers in workplace cafeterias: an exploratory analysis of stated behavior and actual behavior. Resour. Conserv. Recycl. 136, 88–94. https://doi.org/10.1016/j.resconrec.2018.04.015.

Sigala, E.G., Gerwin, P., Chroni, C., Abeliotis, K., Strotmann, C., Lasaridi, K., 2025. Reducing food waste in the HORECA sector using AI-based waste-tracking devices. Waste Manag. 198, 77–86. https://doi.org/10.1016/j.wasman.2025.02.044.

Sundin, N., Malefors, C., Strotmann, C., Orth, D., Kaltenbrunner, K., Obersteiner, G., Scherhaufer, S., Sjölund, A., Persson Osowski, C., Strid, I., Eriksson, M., 2024. Sustainability assessment of educational approaches as food waste prevention measures in school catering. J. Clean. Prod. 481, 144196. https://doi.org/10.1016/j.jclepro.2024.144196.

United Nations, 2015. Transforming our world: the 2030 Agenda for Sustainable Development. https://undocs.org/en/A/RES/70/1.

van der Werf, P., Seabrook, J.A., Gilliland, J.A., 2020. Food for thought: comparing self-reported versus curbside measurements of household food wasting behavior and the predictive capacity of behavioral determinants. Waste Manag. 101, 18–27. https://doi.org/10.1016/j.wasman.2019.09.032.

van Herpen, E., van Geffen, L., Nijenhuis-de Vries, M., Holthuysen, N., van der Lans, I., Quested, T., 2019. A validated survey to measure household food waste. MethodsX 6, 2767–2775. https://doi.org/10.1016/j.mex.2019.10.029.

Vardopoulos, I., Abeliotis, K., Lasaridi, K., 2024. A systematic informetric Analysis and literature Review of food waste quantification studies in the food service industry. Sustainability 17 (1), 103. https://doi.org/10.3390/su17010103.

Winnow. (2025). *Your food waste problem is solved*. https://www.winnowsolutions.com/.

WRAP. (2009). *Household food and drink waste in the UK* . https://www.wrap.ngo/sites/default/files/2020-12/Household-Food-and-Drink-Waste-in-the-UK-2009.pdf.

WRAP, 2013. Household food and drink waste in the United Kingdom 2012. https://open.barnet.gov.uk/download/e67p2/1lc/D5%20WRAP%20report%20Household%20food%20and%20drink%20waste%20uk%202012.pdf.