Integrated multiomic wastewater-based epidemiology can elucidate population-level dietary behaviour and inform public health nutrition assessments

Devin A. Bowes [1,2,3], Erin M. Driver [1], Sonja Savic[1], Qiwen Cheng [2], Corrie M. Whisner[2,4], Rosa Krajmalnik-Brown [2,5] & Rolf U. Halden [1,5,6,7]

Population-level nutritional assessments often rely on self-reported data, which increases the risk of recall bias. Here, we demonstrate that wastewater-based epidemiology can be used for near real-time population dietary assessments. Neighbourhood-level, untreated wastewater samples were collected monthly from within an urban population in the south-western United States from August 2017 to July 2019. Using liquid chromatography–tandem mass spectrometry, we identify recurring seasonal dynamics in phytoestrogen consumption, including dietary changes linked to the winter holiday season. Using 16S ribosomal RNA gene amplicon sequencing, we demonstrated the feasibility of detecting sewage-derived human gut bacterial taxa involved in phytoestrogen metabolism, including *Bifidobacterium*, *Blautia* and *Romboutsia*. Combined metabolomic and genomic wastewater analysis can inform nutritional assessments at population scale, indicating wastewater-based epidemiology as a promising tool for actionable and cost-effective data collection to support public health nutrition.

The impact of dietary behaviour on chronic disease prevalence is undisputed, yet gaining a real-time understanding of this diet–disease relationship remains a challenge[1]. Validated survey-based methods (for example, 24 h recall and food frequency questionnaires) for population-level nutritional assessments, such as part of the National Health and Nutrition Examination Survey, are widely implemented, but are often susceptible to recall bias[2–4]. Objectives have been set forth as part of the Healthy People 2030 Initiative by the United States Department of Health and Human Services to increase consumption of unprocessed whole foods through

evidence-based tactics, such as combining nutritional counselling with workplace or school nutrition policies; however, tools supportive of targeted educational interventions in public health nutrition are still lacking[5].

A plant-based diet generally consists of consuming vegetables, fruits, nuts and seeds, with minimal to no consumption of animal-based foods. This type of diet has been shown to prevent, minimize or even reverse symptoms of certain chronic diseases, including type 2 diabetes and cardiovascular disease, which has motivated recent campaigns to promote plant-based dietary behaviour in the United States[6].

[1]The Biodesign Institute Center for Environmental Health Engineering, Arizona State University, Tempe, AZ, USA. [2]The Biodesign Institute Center for Health Through Microbiomes, Arizona State University, Tempe, AZ, USA. [3]School for Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, AZ, USA. [4]College of Health Solutions, Arizona State University, Phoenix, AZ, USA. [5]School for Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA. [6]OneWaterOneHealth, The Arizona State University Foundation, The Biodesign Institute, Arizona State University, Tempe, AZ, USA. [7]Global Futures Laboratory, Arizona State University, Tempe, AZ, USA. ✉e-mail: rolf.halden@asu.edu

Naturally occurring, plant-derived chemicals known as phytoestrogens can serve as biomarkers indicative of a plant-based diet as they are consumed through a wide variety of foods, including cruciferous vegetables, berries, nuts, flaxseed, soybeans, oat and wheat[6–9]. Two major classes of phytoestrogens commonly consumed in the United States include isoflavones (soy-based foods and food products), as well as lignans, which are much more ubiquitous. These two classes have been investigated in individualized human studies utilizing targeted metabolomic techniques to measure the phytoestrogen chemicals genistein, daidzein, enterolactone and equol, coupled with genomics to assess human gut microbial interactions in response to consumption[10,11]. Differences in geographic region or demographics may contribute to varied rates of phytoestrogen consumption, such as in Asian populations where intake is considerably higher (25–50 mg per day) than the United States (1–3 mg per day); however, prevalence of diseases in Asia, such as breast cancer, which is in part linked to phytoestrogen consumption, remains low[12,13]. These findings could be attributed to differences in human gut microbial communities that produce phytoestrogen-derived metabolites, including lignan- and isoflavone-transformed gut microbial products enterolactone and equol, which have been linked to offering cardiovascular- and cancer-protective benefits for the human host[14,15]. Human gut bacterial genera reported to participate in these enterolactone- and equol-producing pathways include *Ruminococcus*, *Slackia* and *Bacteroides*, rendering these and related commensals prime investigative targets when studying plant-based diets[16–18]. These unique differences in consumption patterns and relationship to disease warrant the need for alternative methods that have the ability to determine population-level dietary information to assess the need for and effectiveness of dietary interventions[19].

Wastewater-based epidemiology (WBE) is a rapidly evolving scientific discipline that leverages the analysis of biochemical signatures detectable in community wastewater conveying composited human excreta (that is, urine and stool) to offer insights into population-level health, behaviour, exposure and activity[20]. Historically employed primarily to monitor substance use across cities[21,22], WBE has gained wide-scale attention and acceptance as a companion tool to clinical testing for monitoring the coronavirus disease 2019 (COVID-19) pandemic at multiple spatial scales (that is, country, city, neighbourhood and/or building level)[23,24]. The success of implementing WBE for viral monitoring with rapid dissemination of actionable data has provided a powerful impetus to leverage this methodology for other public health fields. Use of WBE for elucidating dietary behaviour has been repeatedly proposed as a logical next step; however, only few experimental reports exist so far[25,26]. The complex relationship between dietary intake and subsequent human metabolism warrants multi-pronged analyses for the successful deployment of WBE for advanced dietary population health assessment.

Here, we report on a hypothesized proof-of-concept investigation of plant-based dietary behaviour demonstrating the use of WBE in public health nutrition assessment through targeted metabolomic and genomic analyses. Study goals were to: (1) investigate longitudinal trends of chemical indicators of isoflavone (genistein and daidzein) and lignan (enterolactone) consumption at the subsewershed level, (2) evaluate the relationship between daidzein consumption and associated metabolite equol to enhance insights into isoflavone intake, (3) assess human gut microbial composition of biologically relevant taxa known to play a role in phytoestrogen metabolism, (4) explore relationships between simultaneously consumed foods for a judicious data interpretation and (5) propose inclusion of WBE into public health dietary assessment.

Results

Using a multiomic approach to WBE, multiple chemical and biological biomarkers of phytoestrogen consumption were successfully detected and analysed in community wastewater to indicate plant-based dietary behaviour. Resultant near-real-time population-level data were then used to propose integrating WBE into the public health nutrition framework.

Study location and samples analysed

This study took place in a small (<10,000 population) subsewer catchment within a large urban city located immediately south-east of a major public university. This wastewater catchment is predominantly residential, comprising a mix of single-family homes, condominiums, apartment complexes and off-campus student housing. Demographic estimates based on self-reported data for this catchment area are as follows: white non-Hispanic (40%), Asian (25%), Hispanic or Latino (21%), African American (7%) and American Indian/Alaska Natives (4%). Daily 24 h composited wastewater samples were collected for seven consecutive days per month from August 2017 to July 2019, resulting in a total of 156 samples collected, processed and analysed for chemical and biological indicators of phytoestrogen consumption.

Per capita chemical measurements in community wastewater

In the first year of this neighbourhood-level monitoring study (August 2017 to July 2018), total phytoestrogen consumption patterns (defined as the sum of genistein, daidzein and enterolactone) displayed a distinct increase at the start of each new year (January), with a gradual subsequent decline through March. In the following year (August 2018 to July 2019), this pattern was repeated, with average overall per capita consumption in year 2 (5.0 ± 2.3 mg per day) greater than year 1 (4.1 ± 2.2 mg per day) (Fig. 1a and Supplementary Fig. 1). While both yearly measurements for total phytoestrogen consumption were higher than the estimated average for the United States from traditional methods (1–3 mg per day per capita), individual isoflavone and lignan measurements of consumption fell within this expected range (mg per day per capita): isoflavones 2.7 ± 1.5 (year 1) and 3.5 ± 1.6 (year 2); lignan 1.6 ± 0.9 (year 1) and 2.1 ± 1.0 (year 2) (Fig. 1e,f). Measured daily concentrations in raw wastewater, calculated mass loads and estimated average per capita excretion are shown for genistein, daidzein and enterolactone in Supplementary Figs. 2 and 3a. Statistical comparisons revealed significant differences ($P \leq 0.01$) between years 1 and 2 (Fig. 1b), as well as seasonal changes between fall and winter, and spring and summer (Fig. 1d). Phytoestrogen consumption was slightly lower overall on weekends versus weekdays; however, these differences were subtle ($P > 0.05$) (Fig. 1c). Results of all statistical comparisons for changes in dietary behaviour are listed in Supplementary Table 4. An assessment of the potential for temperature-dependent in-sewer degradation of analytical targets suggested weak susceptibility for genistein ($\rho = 0.22$), daidzein ($\rho = 0.16$) and enterolactone ($\rho = 0.23$) (Supplementary Fig. 4), implying that the reported wastewater-derived, phytoestrogen-metabolite loads are a reflection of human excretion activity and, by extension, human dietary behaviour.

Assessing precursor transformation interactions

The isoflavones investigated in this study, daidzein and genistein, are both precursor (that is, human ingested) compounds, and thus it was necessary to test the efficacy of using these as indicators of phytoestrogen consumption by incorporating an associated human excreted metabolite. Equol, a human gut microbial metabolite of the isoflavone daidzein, was chosen for this purpose and integrated into the study starting from January to July 2019 ($n = 49$ observations). Measured daily concentrations, calculated mass loads and per-capita excretion are listed in Supplementary Fig. 3b. Trends observed here for equol aligned with those of the other phytoestrogens investigated in this study, with elevated consumption at the beginning of the year and a subsequent decline after March. Statistical analysis revealed a strong positive correlation between daidzein and equol ($\rho = 0.84$), with a precursor-to-metabolite ratio of 0.2 (Fig. 2), relatively in line
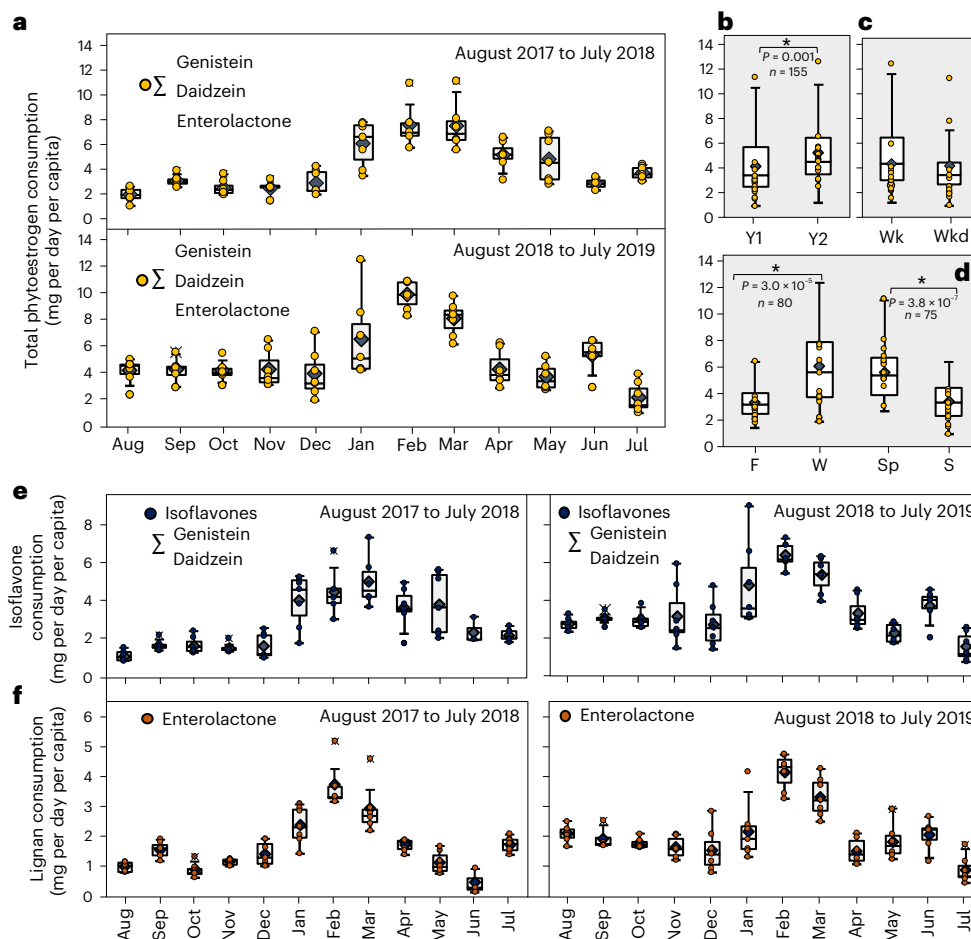
Fig. 1 | Total per capita phytoestrogen consumption (mg per day) from August 2017 to July 2019. a, Total phytoestrogen consumption (sum of genistein, daidzein and enterolactone) shown per month from August 2017 to July 2018 (top) and August 2018 to July 2019 (bottom). **b**–**d**, Statistical comparisons by year (Y) (**b**), weekday (Wk) versus weekend (Wkd) (no statistical significance) (**c**) and season: fall (F) and winter (W); spring (Sp) and summer (S) (**d**). **e**, Isoflavone consumption shown as the sum of genistein and daidzein per month from August 2017 to July 2018 (left) and August 2018 to July 2019 (right).

f, Lignan consumption indicated by the production of enterolactone per month from August 2017 to July 2018 (left) and August 2018 to July 2019 (right). All box plots: centre lines indicate the median; whiskers indicate the range (minimum and maximum); limits are 25th and 75th quartiles. Black diamond represents mean with individual dots representing daily measurements. Statistical significance tested using Mann–Whitney U non-parametric test of variability with BH correction for false discovery (0.05) (*$P \leq 0.01$).

with conventional estimates for equol production (30% of people in Westernized societies)[27]. Equol production reported by WBE can therefore be reported in two different ways: as an average per capita result and as the output per equol producer depending on the study population (Fig. 2). Once adjusted for the population-specific producer to non-producer ratio, computed normalized values are directly comparable and transferable between studies and different demographics.

Population-level dietary human gut microbial interactions
As a proof of concept for investigating human gut microbial interactions at the population scale, a subset of samples (n = 84; January to December 2018) was tested to determine if the bacterial genera identified to be involved in phytoestrogen metabolism and production of enterolactone and equol from individualized studies (Supplementary Table 6) are also observable at the community level via analysis of composited human waste extant in municipal wastewater. Genus-level compositional analysis revealed the presence of specific bacterial taxa known to be involved in phytoestrogen metabolism within the human gut, such as *Bifidobacterium*, *Blautia* and *Romboutsia*, were detectable in community sewage (Fig. 3a). A detection frequency of 100% in all samples was found for *Ruminococcus*, *Blautia*, *Clostridium*,

Bifidobacterium, *Romboutsia*, *Dorea* and *Subdoligranulum*, followed by *Intestinibacter* (66%), *Eubacterium* (58%), *Bacteroides* (50%), *Prevotella* (25%), *Senegalimassilia* and *Roseburia* (16%) and *Slackia* (8%). Measurements of 16S ribosomal RNA gene copies per litre provided an indication of bacterial concentration in each wastewater sample, ranging from 4.3×10^{11} to 2.4×10^{12} (Fig. 3b). Average semi-quantitative abundances of the number of 16S rRNA genes belonging to select enterolactone- and equol-producing genera were calculated, with resultant values ranging from 9.0×10^6 (*Eubacterium*; June) to 2.6×10^{11} (*Bifidobacterium*; October). Estimated average monthly phytoestrogen consumption, determined from chemical biomarker measurements, are shown alongside average semi-quantitative abundances of population-level phytoestrogen-metabolizing taxa to enable a delineation of trends. Elevated trends in both sets of data were observed at the beginning of the year and again towards the end of the summer heading into the fall (autumn), displaying a weak and non-statistically significant association (ρ = 0.24 and P > 0.05), which was, however, intriguing (Fig. 3c).

Evaluating influences from diet variety
In a typical human diet, multiple types of foods and food products are consumed simultaneously, including alcoholic beverages during

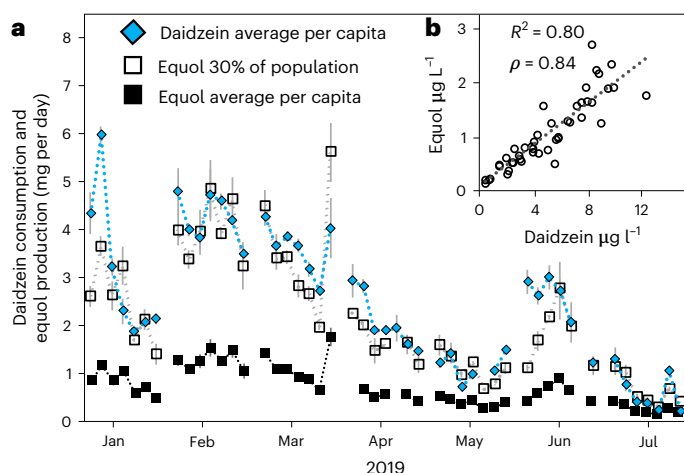**Fig. 2 | Evaluating the relationship between daidzein and equol. a**, Population-scale daidzein consumption (light-blue diamonds) and metabolite equol per active equol shedder (30% of population; shown as white squares) observed in an urban population of <10,000 people using wastewater analysis. Data are shown as mean values between duplicates ($n = 2$); error bars represent standard deviation. **b**, Paired observations ($n = 49$) of concentrations of daidzein and equol in community wastewater (μg l^{-1}) showed a strong linear relationship ($R^2 = 0.80$) and strong positive association ($\rho = 0.84$) using Spearman's rank-order non-parametric correlation test ($n = 93$).
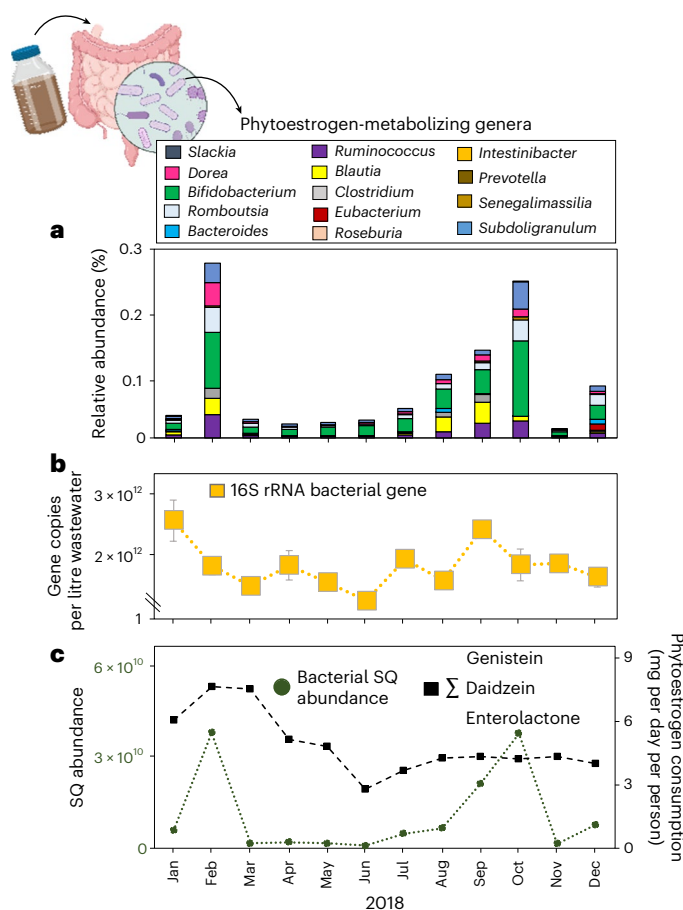

Fig. 3 | Microbial composition of wastewater collected in this study between January and December 2018. a, Relative abundance of select bacterial genera involved in phytoestrogen metabolism within the human gut. **b**, Copy number of bacterial 16S rRNA genes per litre determined using qPCR. Error bars represent standard deviation between duplicates ($n = 2$). **c**, Average semi-quantitative (SQ) abundances (gene copies per litre) of the same phytoestrogen-metabolizing genera shown with total phytoestrogen consumption (mg per day per person) per month.

certain mealtimes or celebrations. A sharp increase in equol production occurred in March 2019 (μg per day per person); 207 ± 7.1 from Saturday 16 March to 594 ± 65.1 on Sunday 17 March, the St Patrick's Day holiday; prompting further investigation into potential interactions with other human consumption behaviours during this particular holiday (alcohol), and its influence on production of equol at the population scale (Fig. 4). Calculated population normalized mass loads (mg per day per person) of equol and ethyl sulfate, a metabolite of alcohol, were compared from January to April 2019 to assess trends, capture holiday behavioural anomalies and seek out concordance in trends of the various analytes. The highest values for ethyl sulfate within this timeframe occurred over the St Patrick's Day holiday: Saturday 16 March (13.6 ± 1.5 mg per day per person) and Sunday 17 March (13.5 ± 1.5 mg per day per person), indicating a pointed increase in alcohol consumption that strongly correlates with the increase observed in equol production ($\rho = 0.89$) (Fig. 4).

Discussion

This study yielded corresponding wastewater-derived datasets on phytoestrogen intake as a proxy for a plant-based diet and the gut microbiome of local human populations. For all compounds examined, there was a considerable increase in consumption starting in January, suggestive of temporary lifestyle changes associated with individual New Years' resolutions of the study population, such as to increase intake of plant-based foods to achieve optimal health[28] (Fig. 1). Additionally, while the notable spike in equol production during the St Patrick's Day holiday period in March 2019 could be from collective participation in consumption of traditional Irish foods, such as colcannon, fish and chips or corned beef and cabbage, which all contain varied amounts of phytoestrogens, it may more likely be due to an increase in acute alcohol consumption[29]. Interestingly, in a study comparing equol producers and non-equol producers, those who reported greater alcohol consumption were more likely to be stronger equol producers; however, the mechanism behind this is still unclear[30]. This unique relationship was tested here at the population scale (Fig. 4), exhibiting strong positive correlations between the alcohol metabolite, ethyl sulfate[31], and equol, suggesting that alcohol

consumption may serve as a driving factor behind equol production when consumed in high amounts and within a condensed timeframe; however, more studies are needed to confirm this phenomenon.

As mentioned, it is estimated that the average consumption rate of phytoestrogens in a Western society range between 1 and 3 mg per day per person; however, consumption can vary depending on life stage, dietary preferences, demographics and/or regular access to healthy fruits and vegetables[32,33]. Thus, this reported estimate of the United States average may not be entirely applicable to this study sewershed with a distinct demographic profile. For example, demographics across the nation consist of 60% white non-Hispanic, 5% Asian, 18% Hispanic or Latino, 13% Black and 1% American Indian/Alaska Native, whereas in this study, white non-Hispanic is 40%, with 25% Asian, 21% Hispanic or Latino, 7% Black and 4% American Indian/Alaska Native[34]. This difference in demographic distribution could result in the disparity between the estimated United States consumption rate and what was measured herein using WBE at the subsewershed level. Measured wastewater-derived isoflavone consumption was closer to the 3 mg per day ceiling than lignans (Fig. 1e), which could be attributed to the greater percentage of Asian residents within this catchment who reportedly consume a greater amount of isoflavones than in Westernized cultures[35]. These differences further demonstrate
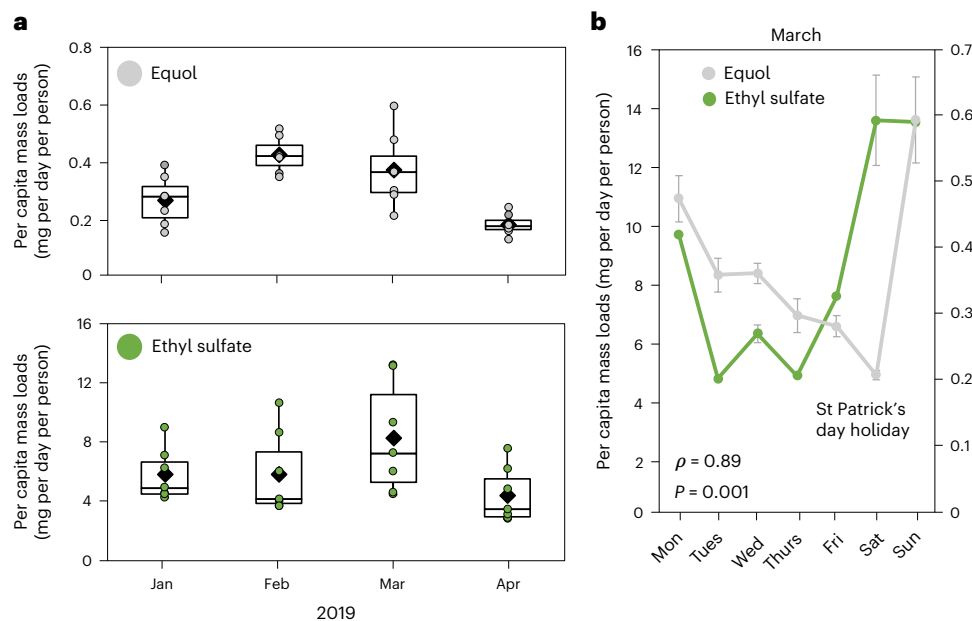
Fig. 4 | Relationship between phytoestrogen and alcohol consumption from January to April 2019. a, Equol production from isoflavone (top) and ethyl sulfate consumption (bottom) from January to April 2019. Box plots: centre lines indicate the median; whiskers indicate the range (minimum and maximum); limits are 25th and 75th quartiles. Black diamond represents mean with individual dots representing daily measurements. **b**, March 2019 displayed Monday to Sunday showing increases over the St Patrick's Day holiday weekend. Statistical comparisons and significance ($P \le 0.01$) suggest strong positive relationship between equol and ethyl sulfate ($\rho = 0.89$ and $P = 0.001$) confirmed by Spearman's rank-order non-parametric test and Mann–Whitney U non-parametric test of variability with BH correction for false discovery (0.05). Error bars represent standard deviation between duplicates ($n = 2$).

the importance of performing more contextually relevant and culturally competent measurements that reflect the population sample under study.

It is evident that human gut microbial interactions that occur in response to plant-based food consumption may hold substantial benefits for the human host. Evidence suggests an increase in regular consumption of plant-based foods can promote gut microbial growth amenable for production of enterolactone and equol, which offer protection against certain nutrition-related chronic diseases[16]. While it is currently understood that approximately 30% of Westernized regions are equol producers, a contextualized investigation could provide deeper insight into this statistic, such as the daidzein-to-equol (precursor-to-metabolite) ratio of 0.2 (20% equol producers) observed in this study. Specifically for enterolactone, studies relying on urinary analysis of individuals reported a noted increase in production to be associated with protective properties against cancer and cardiovascular disease, with use of antibiotics shown to decrease production, indicating the measured values are driven by gut microbes[36]. Studies have also found positive associations between intake of isoflavone-rich foods and subsequent urinary elevations in equol production[11,37], affirming the role of diet in promoting growth of health-promoting commensals[38]. Thus, following a plant-based diet has been shown to strongly correlate with increased production of particular bacterial genera responsible for producing enterolactone and equol in individuals, including *Ruminococcus, Eubacterium, Slackia, Bacteroides* and *Bifidobacterium*[11,39]. These were detected and measured here in community wastewater, with similar trends when compared with phytoestrogen consumption (Fig. 3c). While microbial communities have been assessed in community sewage to understand overall human health[40], this approach to investigating dietary behaviour using WBE has yet to be fully explored, and thus serves as a foundational framework for future analysis of population nutritional assessment (Fig. 5).

The chemical measurements reported herein are slightly higher than those found in a prior WBE study confirming the viability of measuring genistein, daidzein and enterolactone in monthly composites of raw wastewater entering a local wastewater treatment plant to assess diet[25]. However, this previous study collected one single 24 h composite sample per month from a wastewater treatment plant, followed by shipment before sample processing and analysis. Shorter hydraulic residence times, contextual ascertainment of data from subsewershed monitoring, multiple analytical methods and targets employed, and a rapid analysis of samples without delay probably contributed to the comparatively stronger and more detailed dataset observed in the present work. Finally, a cost analysis was conducted from this work, which yielded per-person costs of <US$1 with existing startup equipment, or ~US$2 if capital expenses are needed (Supplementary Table 5). This is considerably less than the estimated average cost for traditional nutritional assessment surveys[41].

This study had limitations that should be noted. Nationwide average ranges from traditional population-level survey methods for phytoestrogen consumption were used for benchmark validation of WBE results; however, these datasets may not be a suitable comparison to the specific catchment area investigated as mentioned above. Additionally, samples in this study were collected according to compliance monitoring pre-determined by the partnering municipality as previously explained. Data obtained were utilized to assess overall yearly trends based on these daily measurements and aggregated per month for ease of visualization (approximately 84 samples per year), which has been indicated in prior work to be suitable for WBE[42]. The minimum number of days needed to represent the year is approximately 57 according to Cochran's formula (10% precision, 90% confidence interval), thus, the number of sample collection days here are considered acceptable for this purpose (supplementary equation (1)) (ref. [43]). While an alternative approach to sampling (for example, weekly) may have offered different insights, it is important to consider the resources available and weigh against the potential benefits when conducting a WBE study. Further, while trends could be garnered between microbial taxa and measured phytoestrogens, genus-level reporting may be limiting for population-level analysis, therefore more detailed investigation incorporating metagenomics is warranted
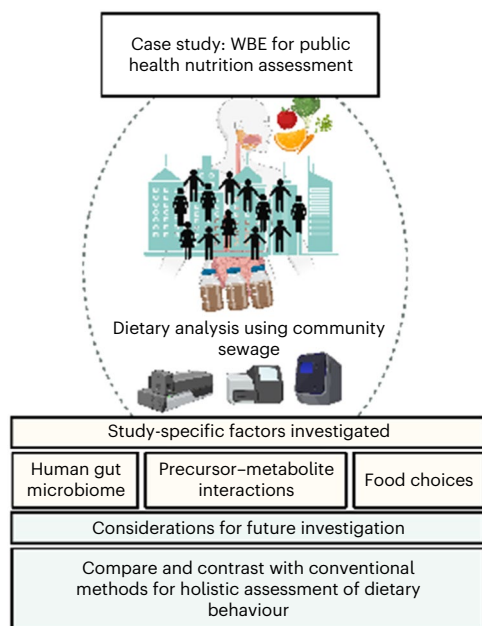
Fig. 5 | Case study-derived conceptual framework for integrating WBE into public health nutrition.

in future work. Additionally, method detection limits and analyte recovery experiments were conducted at the start of this 2 year study. While instrument performance was continuously assessed by analysing standard curves made monthly, and use of isotopically labelled internal standards can assist in evaluating variability over time, it is acknowledged that matrix changes may have occurred over the study duration. Finally, use of subsewershed monitoring served to reduce potential in-sewer loss of analytes from the point of human excretion to the sampling location, and while temperature-influenced degradation investigated herein exhibited weak susceptibility (Supplementary Fig. 4), some analyte loss may still have occurred.

Overall, this study highlights the utility, importance and cost advantages of integrating multiomic WBE into population nutritional assessments. Here we report a successful demonstration of longitudinal investigation showcasing continuous signal detectability, considerable dynamic range and joint trending of wastewater-borne biomarkers that are modulated as a function of human diet and behaviour. Major findings of this work include: (1) multi-parameter community-scale datasets showing the interplay between human behaviour (for example, alcohol consumption on holidays and short-term dietary changes) and dietary indicators detectable in composited municipal wastewater; (2) association between wastewater-borne levels of phytoestrogens and related gut microbial composition; (3) the introduction of daidzein metabolite, equol, to offer a deeper interpretation of isoflavone consumption measured in wastewater and (4) potential relationships between simultaneously consumed foods that may serve as driving factors for subsequent production of associated metabolites. The data reported here suggest that the integration of WBE could offer a holistic model for public health nutritional assessment by providing contextual, non-invasive and actionable data for community stakeholders to inform on nutritional status, the need for and appropriate design of targeted strategies, and the success of such interventions chosen at differing scales (for example, school lunches, university campuses and urban populations).

Methods
Chemicals and reagents
Native analytical standards of genistein, daidzein and enterolactone were purchased from Sigma-Aldrich. Isotopically labelled analytical

standard genistein-*d4* was purchased from Cayman Chemical. Native analytical standard of equol was purchased from Santa Cruz Biotechnology. Native analytical standard of ethyl sulfate (EtS) as well as its isotopically labelled standard, sodium ethyl sulfate-*d5* (EtS-*d5*) were purchased from Toronto Research Chemicals. Liquid chromatography-grade (LC-grade) water, methanol and acetone were obtained from Fisher Scientific, and LC-grade formic acid was purchased from Fluka Chemical Corp. Stock standard solutions were prepared in LC-grade methanol and stored at −20 °C, with the exception of EtS and EtS-*d5*, which were stored in water at 4 °C.

Sample collection and transport
The wastewater collection system in this sewer catchment is separate from the stormwater collection system, and drains south-west, with an average sewage retention time of approximately 20 min. Recorded measurements of wastewater temperatures unique to this catchment ranged between 20 °C and 30 °C depending on the time of year. Time-weighted, 24 h composite wastewater samples were collected daily from within the sewer collection system (that is, neighbourhood level) over seven consecutive days per month between August 2017 and July 2019 (n = 156), allowing for a robust indication of consumption trends over time given the number of observations per year[42]. An Avalanche automated refrigerated sampler (Teledyne, ISCO) was set to collect 60–100 ml of raw wastewater from within the sewer collection system every 15 min. Each 24 h composited sample was adequately mixed and transferred to two-litre high-density polyethylene bottles and immediately placed on ice to prepare for transport and same-day processing. Wastewater flow measurements were monitored by ISCO LaserFlow flow metres (Teledyne, ISCO), which monitored and recorded flow on-site at 2 min intervals. Flow measurements were obtained through FlowLink online software (v. 5.1) (Teledyne, ISCO).

Sample processing and analysis
Methods employed for chemical processing and analysis have been previously reported[25,31]. Briefly, duplicated 200 ml aliquots of each raw influent wastewater sample were spiked with 20 ng isotopically labelled internal standard and subsequently arranged on a Dionex Autotrace 280 Solid-Phase Extraction Instrument (Thermo Scientific) using Oasis Hydrophilic-Lipophilic Balance cartridges (Waters). Method blanks (deionized water) were extracted and analysed alongside each set of samples to determine potential contamination and assess method performance over time. Analytes of interest were then extracted using a 1:1 (vol/vol) methanol and acetone solution with 0.5% formic acid until a final volume of 4 ml was achieved, then stored at −20 °C until further analysis. For EtS, 10 ml of raw wastewater sample was aliquoted in 15 ml conical tubes and centrifuged at 4,000g for 10 min. Next, 500 μl of the supernatant were added to 490 μl of water along with 1 ng labelled standard (EtS-*d5*) to achieve a final extract volume of 1 ml ready for analysis by liquid chromatography–tandem mass spectrometry (LC–MS/MS).

For LC–MS/MS sample extract preparation (except EtS), 200 μl of each organic extract was aliquoted into glass amber vials and dried down entirely under nitrogen. Extracts were reconstituted first with 100 μl of LC–MS-grade methanol, followed by 100 μl of LC–MS-grade water, then lightly vortexed. Finalized extracts were analysed for targeted analytes using a Shimadzu Prominence 2100 high performance liquid chromatographer (Marlborough) paired to an AB Sciex API 4000 triple quadrupole mass spectrometer (Applied Biosystems) with electrospray ionization operating in negative mode. Analyte identification was achieved using compound-specific retention times and ion transitions from multiple reaction monitoring. Chromatographic separation was attained using a Symmetry C_8 column (4.6 × 150 mm and 3.5 μm particle size), followed by a Symmetry VanGuard Cartridge (3.9 × 5 mm and 3.5 μm particle size) (Waters) (Supplementary Table 1).

Data analysis

LC–MS/MS data were acquired with Analyst 1.5 software (Applied Biosystems), where concentrations were calculated using isotope dilution. Calculated concentrations (µg l⁻¹) were converted to mass loads (g day⁻¹) using flow data accessed with FlowLink software (approximately 2,165,000 litres per day from August to May and 1,637,000 litres per day from June to July).

Population-normalized mass loads and subsequent consumption rates were produced using population estimates (approximately 9,800 from August to May and 6,900 from June to July) and estimated excretion values listed in Supplementary Table 3. Per capita daily genistein consumption (GC) was calculated using the following equation:

$$GC = \frac{C_G \times Q_{Tot}}{CF_G \times Pop} \qquad (1)$$

where C_G is the measured genistein concentration, Q_{Tot} is the total daily volumetric flow rate and CFG is the correction factor (5; 20% excretion). Pop is estimated population. Per capita daily daidzein consumption (DC) was calculated using the following equation:

$$DC = \frac{C_D \times Q_{Tot}}{CF_D \times Pop} \qquad (2)$$

where C_D is the measured daidzein concentration, Q_{Tot} is the total daily volumetric flow rate and CF_D is the correction factor (2.2; 45% excretion). Pop is estimated population. Per capita daily lignan consumption (LC) was calculated using the following equation:

$$LC = \frac{C_{ENT} \times \left(\frac{MW_{LIG}}{MW_{ENT}}\right) Q_{Tot}}{EF_{ENT} \times Pop} \qquad (3)$$

where C_{ENT} is the measured enterolactone concentration, MW_{LIG}/MW_{ENT} is the ratio of molecular weights with the average of four main parent lignans common in the human diet; pinoresinol, lariciresinol, matairesinol and secoisolariciresinol (LIG) and enterolignan metabolite enterolactone (ENT), Q_{Tot} is the total daily volumetric flow rate, EF_{ENT} is the urinary excretion of enterolactone (1.1 mg per day). Pop is estimated population. Per capita daily equol production (EC) was calculated using the following equation:

$$EC = \frac{C_{EQ} \times \left(\frac{MW_{DAD}}{MW_{EQ}}\right) Q_{Tot}}{EF_{EQ} \times Pop} \qquad (4)$$

where C_{EQ} is the measured equol concentration, MW_{DAD}/MW_{EQ} is the ratio of molecular weights of the parent daidzein (DAD) and metabolite equol (EQ), Q_{Tot} is the total daily volumetric flow rate, EF_{EQ} is the urinary excretion of equol (2.7 mg per day). Pop is estimated population. Ethyl sulfate population mass loads followed the same equations as above to represent alcohol use.

The population of contributing individuals within this subsewer catchment were determined using previously reported methods[44]. Briefly, this catchment predominately comprises permanent residents, as estimated by 2010 census block group data. To factor in transient populations, employment data from Maricopa Association of Governments, resident and non-resident employment was examined. The following classifications were used: employees living outside of the city (non-resident and employed) and residents of the city (resident and employed). A portion of residents in this catchment are students attending the nearby university with housing contracts typically for the entire academic year (August to May). These student population estimates were obtained from publicly available campus resident data and estimates using changes in wastewater flow volume to make adjustments at the beginning and end of the academic year.

Statistics and reproducibility

Reported concentrations were determined on the basis of a minimum seven-point standard curve ranging from 0.05 to 2,000 µg l⁻¹, and a minimum coefficient of determination of $r^2 = 0.99$. Precision was expressed as relative percentage difference (RPD) (equation (5)). Instrument blanks were included every six to eight samples to assess analyte carryover, of which none was observed. Method detection limits were determined using methods outlined by the United States Environmental Protection Agency following recommendations by EPA Method 1694 (ref. [45]) with daidzein, genistein and enterolactone previously reported and validated[25]. Average absolute recoveries (before normalization with stable isotope internal standards) for target analytes were determined on the basis of spike-recovery experiments in a peat moss matrix according to previously reported methods[25,46] (Supplementary Table 2).

RPD was calculated using the following equation:

$$RPD\% = ABS\left(\frac{C_{S1} - C_{S2}}{C_{S1} + C_{S2}}/2\right) \times 100 \qquad (5)$$

where C_{S1} and C_{S2} are the measured concentrations in the sample and its associated duplicate.

Statistical analyses were performed using Microsoft Excel 2019, where Mann–Whitney U and Spearman rank-order non-parametric tests were conducted. To control for type I errors and correct for multiple tests, a Benjamini–Hochberg (BH) correction factor was applied (false discovery rate 0.05) using the following equation:

$$BH = \frac{i}{m}Q \qquad (6)$$

where i is the rank assigned to the P value in the array, m is the number of comparisons and Q is the false discovery rate. To define seasonality, seasons were determined according to the National Geographic Society:[47] fall (September, October and November), winter (December, January and February), spring (March, April and May) and summer (June, July and August). Formal randomization is not relevant for this type of study as the inherent nature of WBE randomly collects contributions of the individuals within any particular sewer catchment at any given time. Further, sample collection weekly campaigns (seven consecutive days per month) were pre-determined by the city, by which the researchers needed to comply. This could contribute to further randomizing the collected sample. Thus, no statistical method was used to pre-determine sample size, the experiments were not intentionally randomized and no data were excluded from the analyses. Additionally, formal blinding is not relevant for this type of study as the nature of WBE inherently and effectively blinds the researchers to the contributing population, regardless of geospatial scale. A major benefit of performing this type of work is the anonymous nature of the collected and analysed sample; individuals are unable to be identified. Thus, the resultant data are representative of the sewer catchment served as a whole.

Effect of temperature on in-sewer analyte degradation was assessed through statistical analysis based on real measurements. Reported ambient temperatures (average, minimum and maximum) for each sample collection day throughout this study were recorded. The relationship between recorded ambient temperatures and measured analyte signal in wastewater (µg l⁻¹) for each analyte undergoing long-term monitoring (genistein, daidzein and enterolactone) was assessed using Spearman's rank-order non-parametric tests ($\rho < 0.50$ weak; $\rho > 0.50 < 0.70$ moderately strong; $\rho > 0.70$ strong) (Supplementary Fig. 4).

Microbiome analysis

As a proof of concept for understanding human gut microbial interactions at the population level, a subset of samples ($n = 12$ months)

was allocated for microbiome analyses. Approximately 84 previously frozen raw wastewater samples from 2018 (January–December; seven samples per month) were thawed, mixed and aliquoted into composites representing each month. Next, approximately 50 ml of each sample was loaded onto sterile 0.22 μm polycarbonate membrane filters (47 mm) (EMD Millipore) using a vacuum pump apparatus with in-line filter, discarding the filtrate. Each filter was then aseptically transferred into an individual bead tube. DNA extractions were performed using a QIAGEN DNeasy Power Soil Pro Kit, following the manufacturer's instructions. The resultant DNA (50 μl) was immediately stored at −80 °C until further analysis. A whole-process negative extraction control (deionized water) was incorporated to account for potential contamination.

Bacterial community composition analysis was performed with next-generation sequencing. The 16S rRNA V4 region was polymerase chain reaction (PCR) amplified using the barcoded primer set 515f/806r (ref. [48]), following the protocol by the Earth Microbiome Project (www.earthmicrobiome.org) for library preparation[49]. PCR amplifications for each sample were performed in duplicate, then pooled and quantified using the Accublue High sensitivity dsDNA Quantitation Kit (Biotium). A no/template control was included during the library preparation for extraneous nucleic acid contamination. After PCR, 200 ng of DNA per sample were pooled and cleaned using a QIA quick PCR purification kit (Qiagen). The pool was quantified by Illumina library Quantitation Kit ABI Prism (Kapa Biosystems). The DNA pool was then diluted to a final concentration of 4 nM, then denatured and diluted to a final concentration of 4 pM with 25% of PhiX for quality control. Finally, the DNA library was loaded onto the Illumina MiSeq platform and run using the version 2 module (2 × 250 paired end) following the directions of the manufacturer.

Data sequences were analysed using QIIME 2 (version 2021.2) (ref. [50]) for sequence quality control and feature table construction. The DADA2 plugin[51] was used to filter and merge the forward and reverse reads. Sequences were then mapped to the Silva 138 SSURef NR99 Database[52] (updated 27 August 2020) for microbial community composition analysis. A classifier was trained with the forward and reverse primers used in this study. The generated QIIME 2 files were imported into R (v. 4.1) using the R phyloseq package (version 1.36.0) (ref. [53]), and contaminating amplicon sequence variants (ASVs) were identified and removed with the R decontam package (version 1.12.0) (ref. [54]). ASVs were classified as contaminants if identified by either the frequency or prevalence methods (method = 'either'). The relative abundances of individual genus relative to the abundances of total genera in each sample were visualized using the R ggplot2 package (version 3.3.5) (ref. [55]).

Quantitative PCR (qPCR) was performed using a QuantStudio 3 (Applied Biosystems) instrument to quantify the total 16S rRNA gene as a proxy of bacterial concentration (16S rRNA gene copies per litre wastewater) in each sample. A previously published TaqMan-based assay was employed targeting the 16S rRNA gene using universal primers (BAC1055F 5′-ATGGYTGTCGTCAGCT-3′; BAC1392R 5′-ACGGGCGGTGTGTAC-3′) and probe (BAC1115Probe 5′′-FAM-CAACGAGCGCAACCC-3′-TAMRA) (ref. [56]). All samples and controls were run in triplicate; standard curve (16S rRNA plasmids) ranged from 10^7 to 10^1 copies μl^{-1}. Molecular no-template controls using RNAse/DNAse-free Ultrapure PCR-grade water (Invitrogen) were incorporated into each run.

Semi-quantitative calculations were achieved by multiplying the relative abundance of select genera (%) by the resultant copy numbers of the 16S rRNA gene per litre of wastewater informed by qPCR in each corresponding sample, resulting in the number of 16S rRNA genes belonging to each selected genera under investigation.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data needed to evaluate the conclusions in this paper are present in the main text and/or Supplementary Information. Special requests can be made to the corresponding author.

References

1. Neuhouser, M. L. The importance of healthy dietary patterns in chronic disease prevention. *Nutr. Res.* **70**, 3–6 (2019).
2. 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. Nut. Diet.* **112**, 1134–1137 (2012).
3. Maurer, J. et al. The psychosocial and behavioral characteristics related to energy misreporting. *Nutr. Rev.* **64**, 53–66 (2006).
4. Pfeiffer, C. M., Lacher, D. A., Schleicher, R. L., Johnson, C. L. & Yetley, E. A. Challenges and lessons learned in generating and interpreting NHANES nutritional biomarker data. *Adv. Nutr.* **8**, 290–307 (2017).
5. Nutrition and healthy eating. *US Department of Health and Human Services* https://health.gov/healthypeople/objectives-and-data/browse-objectives/nutrition-and-healthy-eating (2022).
6. Medawar, E., Huhn, S., Villringer, A. & Veronica Witte, A. The effects of plant-based diets on the body and the brain: a systematic review. *Transl. Psychiatry* **9**, 226 (2019).
7. Desmawati, D. & Sulastri, D. Phytoestrogens and their health effect. *Open Access Maced. J. Med. Sci.* **7**, 495–499 (2019).
8. Patisaul, H. B. & Jefferson, W. The pros and cons of phytoestrogens. *Front. Neuroendocrinol.* **31**, 400–419 (2010).
9. Rietjens, I., Louisse, J. & Beekmann, K. The potential health effects of dietary phytoestrogens. *Br. J. Pharmacol.* **174**, 1263–1280 (2017).
10. Franke, A. A., Lai, J. F. & Halm, B. M. Absorption, distribution, metabolism, and excretion of isoflavonoids after soy intake. *Arch. Biochem. Biophys.* **559**, 24–28 (2014).
11. Gaya, P., Medina, M., Sánchez-Jiménez, A. & Landete, J. M. Phytoestrogen metabolism by adult human gut microbiota. *Molecules* **21**, 1034 (2016).
12. He, F.-J. & Chen, J.-Q. Consumption of soybean, soy foods, soy isoflavones and breast cancer incidence: differences between Chinese women and women in Western countries and possible mechanisms. *Food Sci. Hum. Wellness* **2**, 146–161 (2013).
13. Messina, M., Nagata, C. & Wu, A. H. Estimated Asian adult soy protein and isoflavone intakes. *Nutr. Cancer* **55**, 1–12 (2006).
14. McCann, S. E. et al. Enterolignan production in a flaxseed intervention study in postmenopausal US women of African ancestry and European ancestry. *Nutrients* **13**, 919 (2021).
15. Setchell, K. D. et al. S-equol, a potent ligand for estrogen receptor beta, is the exclusive enantiomeric form of the soy isoflavone metabolite produced by human intestinal bacterial flora. *Am. J. Clin. Nutr.* **81**, 1072–1079 (2005).
16. Setchell, K. D. & Clerici, C. Equol: pharmacokinetics and biological actions. *J. Nutr.* **140**, 1363s–1368s (2010).
17. Peterson, J. et al. Dietary lignans: physiology and potential for cardiovascular disease risk reduction. *Nutr. Rev.* **68**, 571–603 (2010).
18. Tomova, A. et al. The effects of vegetarian and vegan diets on gut microbiota. *Front. Nutr.* **6**, 47 (2019).
19. Choi, P. M. et al. Do food and stress biomarkers work for wastewater-based epidemiology? A critical evaluation. *Sci. Total Environ.* **736**, 139654 (2020).
20. Lorenzo, M. & Picó, Y. Wastewater-based epidemiology: current status and future prospects. *Curr. Opin. Environ. Sci. Health* **9**, 77–84 (2019).
21. Baz-Lomba, J. A. et al. Comparison of pharmaceutical, illicit drug, alcohol, nicotine and caffeine levels in wastewater with sale, seizure and consumption data for 8 European cities. *BMC Public Health* **16**, 1035 (2016).

22. Gushgari, A. J., Driver, E. M., Steele, J. C. & Halden, R. U. Tracking narcotics consumption at a Southwestern U.S. university campus by wastewater-based epidemiology. *J. Hazard. Mater.* **359**, 437–444 (2018).

23. Ahmed, W. et al. Detection of SARS-CoV-2 RNA in commercial passenger aircraft and cruise ship wastewater: a surveillance tool for assessing the presence of COVID-19 infected travellers. *J. Travel Med.* **27**, taaa116 (2020).

24. Wright, J., Driver, E. M., Bowes, D. A., Johnston, B. & Halden, R. U. Comparison of high-frequency in-pipe SARS-CoV-2 wastewater-based surveillance to concurrent COVID-19 random clinical testing on a public U.S. university campus. *Sci. Total Environ.* **820**, 152877 (2022).

25. Venkatesan, A. K., Chen, J., Driver, E., Gushgari, A. & Halden, R. U. in *Wastewater-Based Epidemiology: Estimation of Community Consumption of Drugs and Diets ACS Symposium Series* Vol. 1319 (eds Subedi, B., Burgard, D. A. & Loganathan B. G.) 187–198 (American Chemical Society, 2019).

26. Choi, P. M. et al. Social, demographic, and economic correlates of food and chemical consumption measured by wastewater-based epidemiology. *Proc. Natl Acad. Sci. USA* **116**, 21864–21873 (2019).

27. Iino, C. et al. Daidzein intake is associated with equol producing status through an increase in the intestinal bacteria responsible for equol production. *Nutrients* **11**, 433 (2019).

28. Pope, L., Hanks, A. S., Just, D. R. & Wansink, B. New Year's res-illusions: food shopping in the new year competes with healthy intentions. *PLoS ONE* **9**, e110561 (2014).

29. Lefkowitz, E. S., Patrick, M. E., Morgan, N. R., Bezemer, D. H. & Vasilenko, S. A. State Patty's day: college student drinking and local crime increased on a student-constructed holiday. *J. Adolesc. Res.* **27**, 323–350 (2012).

30. Bolca, S. et al. Microbial and dietary factors are associated with the equol producer phenotype in healthy postmenopausal women. *J. Nutr.* **137**, 2242–2246 (2007).

31. Driver, E. M., Gushgari, A., Chen, J. & Halden, R. U. Alcohol, nicotine, and caffeine consumption on a public U.S. university campus determined by wastewater-based epidemiology. *Sci. Total Environ.* **727**, 138492 (2020).

32. Fletcher, R. J. Food sources of phyto-oestrogens and their precursors in Europe. *Br. J. Nutr.* **89**, S39–43 (2003).

33. Lampe, J. W. Isoflavonoid and lignan phytoestrogens as dietary biomarkers. *J. Nutr.* **133**, 956S–964S (2003).

34. United States Census quick facts. *United States Census Bureau* https://www.census.gov/quickfacts/fact/table/US/PST045221#qf-headnote-a (2020).

35. Chen, M. et al. Association between soy isoflavone intake and breast cancer risk for pre- and post-menopausal women: a meta-analysis of epidemiological studies. *PLoS ONE* **9**, e89288 (2014).

36. Kilkkinen, A. et al. Use of oral antimicrobials decreases serum enterolactone concentration. *Am. J. Epidemiol.* **155**, 472–477 (2002).

37. Mayo, B., Vázquez, L. & Flórez, A. B. Equol: a bacterial metabolite from the daidzein isoflavone and its presumed beneficial health effects. *Nutrients* **11**, 2231 (2019).

38. Matthies, A., Loh, G., Blaut, M. & Braune, A. Daidzein and genistein are converted to equol and 5-hydroxy-equol by human intestinal *Slackia isoflavoniconvertens* in gnotobiotic rats. *J. Nutr.* **142**, 40–46 (2012).

39. Clavel, T. et al. Intestinal bacterial communities that produce active estrogen-like compounds enterodiol and enterolactone in humans. *Appl. Environ. Microbiol.* **71**, 6077–6085 (2005).

40. Newton, R. J. et al. Sewage reflects the microbiomes of human populations. *mBio* **6**, e02574 (2015).

41. Knüppel, S., Norman, K. & Boeing, H. Is a single 24-hour dietary recall per person sufficient to estimate the population distribution of usual dietary intake. *J. Nutr.* **149**, 1491–1492 (2019).

42. Ort, C. et al. Challenges of surveying wastewater drug loads of small populations and generalizable aspects on optimizing monitoring design. *Addiction* **109**, 472–481 (2014).

43. Cochran, W. G. *Sampling Techniques* 3rd edn (John Wiley & Sons, 1977).

44. Bowes, D. A. et al. Leveraging an established neighbourhood-level, open access wastewater monitoring network to address public health priorities: a population-based study. *Lancet Microbe* **4**, e29–e37 (2023).

45. US EPA. Definition and procedure for the determination of the method detection limit. https://www.epa.gov/sites/default/files/2016-12/documents/mdl-procedure_rev2_12-13-2016.pdf (2016).

46. Supowit, S. D., Sadaria, A. M., Reyes, E. J. & Halden, R. U. Mass balance of fipronil and total toxicity of fipronil-related compounds in process streams during conventional wastewater and wetland treatment. *Environ. Sci. Technol.* **50**, 1519–1526 (2016).

47. Seasons. *National Geographic Society* https://www.nationalgeographic.org/encyclopedia/season/ (2020).

48. Caporaso, J. G. et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108**, 4516–4522 (2011).

49. Thompson, L. R. et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).

50. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).

51. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).

52. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).

53. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).

54. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).

55. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).

56. Ritalahti, K. M. et al. Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple *Dehalococcoides* strains. *Appl. Environ. Microbiol.* **72**, 2765–2774 (2006).

Acknowledgements

Author contributions

D.A.B.: conceptualization, investigation, sample processing, data generation and analysis, writing—original manuscript draft. E.M.D.: investigation, sample processing, data generation and analysis, writing—manuscript review and editing. S.S.: sample processing and analysis, writing—manuscript review and editing. Q.C.: data analysis,

writing—manuscript review and editing. C.M.W. and R.K.-B.: oversight, writing—manuscript review and editing. R.U.H.: conception, oversight, writing—manuscript review and editing

Competing interests

E.M.D. is a managing member of AquaVitas, LLC, a company working in the field of WBE. R.U.H. is a managing member of AquaVitas LLC and of the non-profit project OneWaterOneHealth of the Arizona State University Foundation, with both entities being involved in the same intellectual space. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43016-023-00717-w.

Correspondence and requests for materials should be addressed to Rolf U. Halden.

Corresponding author(s): Rolf U. Halden, PhD, PE

Last updated by author(s): 02/05/2023

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	LC-MS/MS data were acquired with Analyst 1.5 software (Applied Biosystems). Quantitative polymerase chain reaction data (qPCR) were collected using QuantStudio 3 Design & Analysis software (v1.5.2). Data sequences were analyzed using QIIME2 (v2021.2) with DADA2 plugin and mapped to the Silva 138.1 SSURef NR99 Database updated 08/27/2020. Flow measurements were obtained through FlowLink online software (v. 5.1) (Teledyne, ISCO, Lincoln, NE, USA).
Data analysis	Statistical analyses were performed using Microsoft Excel version 2019 where Mann Whitney U Tests and Spearmans rank-order nonparametric tests were conducted with Benjamini-Hochberg correction (FDR = 0.05) to assess consumption patterns. Generated QIIME2 data files were imported into R (v 4.1) using the R phyloseq package (version 1.36.0) and decontam package (version 1.12.0). Relative abundances of total genera in each sample were visualized using R ggplot2 package (v3.3.5).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are presented in the supplementary information (Tables 1-6, Figures 1-4, Equation 1). Special requests can be made to the corresponding author upon reasonable request. Silva 138.1 SSURef NR99 Database (https://www.arb-silva.de/documentation/release-138/).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences   ☐ Behavioural & social sciences   ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study collected untreated wastewater samples from within the sewer infrastructure in a sub-sewer catchment within the City of Tempe, Arizona for a total duration of two years (August 2017 through July 2019) and tested for chemical and biological indicators of a plant-based diet in order to assess population-level dietary behavior. Chemical indicators of a plant-based diet measured include the phytoestrogens genistein, daidzein, enterolactone, and equol. Biological indicators include assessment of microbial communities unique to the human gut microbiome and detectable in a community wastewater sample responsible for metabolizing the aforementioned chemicals once consumed. This is the first to report a multiomics approach (community metabolomic and genomic applications) to performing a population-level dietary assessment using wastewater-based epidemiology, and is proposed herein as a viable tool to supplement existing methods for a comprehensive view of population nutritional habits.
Research sample	The collected sample is an untreated, 24-hour composited community wastewater sample collected daily for seven consecutive days per month (7 separate samples collected each month analyzed individually). This study catchment serves approximately 10,000 population and is 100% connected to the municipal wastewater system, which is separate from the city's stormwater system. Population of contributing individuals to the wastewater sample was evaluated utilizing a validated method which takes into account residential population and transient populations leveraging information obtained from employment and flow data resources.
Sampling strategy	Untreated, 24-hour composited, community wastewater samples were collected daily for seven consecutive days each month from August 2017 through July 2019. Time-weighted automated samplers were set to collect an aliquot of wastewater from within the sewer infrastructure every 15 minutes for a total of 24-hours each day. Each aliquot is collected into a large 10L vessel. The resultant 24-hour daily composited sample is well mixed, poured, and transported to the laboratory for same-day processing. This was repeated for each collection day per month (n=7). Samples were processed and subsequently analyzed individually.
Data collection	Upon arrival to the laboratory, the sample was recorded (date, location) and underwent sample processing (solid-phase extraction) same-day after spiking with isotopically-labeled internal standards of the selected analytes. Post-extraction, samples were subject to gravity-drip elutions, dried down under nitrogen, and the reconstituted to prepare for LC-MS/MS analysis. Other data (flow, ambient temperature on collection day, wastewater temperature on collection day, etc.) were collected and recorded.
Timing and spatial scale	Collected wastewater samples are twenty-four (24) hour composites approximately representing 7:00AM to 6:59AM activity within the community. Information learned from each collected sample is representative of the previous 24-hours of human activity. Each 7-consecutive day sampling campaign that occurred each month represented week and weekend activity. The automated sampler was located in the southwestern region of the sewer catchment. The sub-sewer catchment is located within the eastern region of the City of Tempe located adjacent to a major public university.
Data exclusions	No data were excluded from this study.
Reproducibility	Each collected wastewater sample was processed in duplicate. Precision was expressed as relative percentage difference between the biological replicates; all analytes achieved target precision of <30%. Absolute analyte spike-recovery experiments and detection limits were assessed using peat moss water as a surrogate for wastewater at the beginning of the study, and continuously evaluated throughout the remainder of the study based on instrument performance and normalization to isotopically-labeled internal standards.
Randomization	Formal randomization is not relevant for this type of study as the inherent nature of wastewater-based epidemiology randomly collects contributions of the individuals within any particular sewer catchment at any given time. Further, sample collection weekly campaigns (7 consecutive days per month) were predetermined by the City, by which the researchers needed to comply. This could contribute to further randomizing the collected sample.
Blinding	Formal blinding is not relevant for this type of study as the nature of wastewater-based epidemiology inherently and effectively blinds the researchers to the contributing population, regardless of geospatial scale. A major benefit of performing this type of work is the anonymous nature of the collected and analyzed sample; individuals are unable to be identified. Thus, the resultant data are representative of the sewer catchment served as a whole.

Did the study involve field work?   ☒ Yes   ☐ No

Field work, collection and transport

Field conditions	Raw, untreated wastewater samples were collected from within the sewer infrastructure by trained City personnel in Tempe, Arizona using an above-ground automated sampler enclosed within a permanent steel sampling station. Researchers of this study obtained custody of portions (AKA "splits") of samples at the designated sample collection area. This area is located southwest of the

	catchment served within an off-street parking lot away from traffic in open air. Typical weather conditions for the Greater Phoenix Metropolitan Area applied, with few rainfall events or other inclement weather.
Location	Tempe, Arizona
Access & import/export	The sample collection site could be accessed by car, foot, or other means of transportation with ease. The permanent sampling station could be accessed using lock-and-key with designated staff/researchers as keyholders.
Disturbance	With the exception of the occasional clog in the sampler line common when using an automated sampler to collect raw wastewater, there are no other study disturbances to report.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging