Novel Applications of Wastewater-based Epidemiology for Assessing Population Nutrition, Infectious Disease, and Chronic Illness

by

Devin Bowes

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved February 2022 by the
Graduate Supervisory Committee:

Rolf U. Halden, Co-Chair
Rosa Krajmalnik-Brown, Co-Chair
Otakuye Conroy-Ben
Arvind Varsani
Corrie Whisner

ARIZONA STATE UNIVERSITY

May 2022

ABSTRACT

Traditional public health strategies for assessing human behavior, exposure, and activity are considered resource-exhaustive, time-consuming, and expensive, warranting a need for alternative methods to enhance data acquisition and subsequent interventions. This dissertation critically evaluated the use of wastewater-based epidemiology (WBE) as an inclusive and non-invasive tool for conducting near real-time population health assessments. A rigorous literature review was performed to gauge the current landscape of WBE to monitor for biomarkers indicative of diet, as well as exposure to estrogen-mimicking endocrine disrupting (EED) chemicals via route of ingestion. Wastewater-derived measurements of phytoestrogens from August 2017 through July 2019 (n = 156 samples) in a small sewer catchment revealed seasonal patterns, with highest average per capita consumption rates in January through March of each year (2018: 7.0 ± 2.0 mg d^{-1}; 2019: 8.2 ± 2.3 mg d^{-1}) and statistically significant differences (p = 0.01) between fall and winter (3.4 ± 1.2 vs. 6.1 ± 2.9 mg d^{-1}; $p \leq 0.01$) and spring and summer (5.6 ± 2.1 vs. 3.4 ± 1.5 mg d^{-1}; $p \leq 0.01$). Additional investigations, including a human gut microbial composition analysis of community wastewater, were performed to support a methodological framework for future implementation of WBE to assess population-level dietary behavior. In response to the COVID-19 global pandemic, a high-frequency, high-resolution sample collection approach with public data sharing was implemented throughout the City of Tempe, Arizona, and analyzed for SARS-CoV-2 (E gene) from April 2020 through March 2021 (n = 1,556 samples). Results indicate early warning capability during the first wave (June 2020) compared to newly reported clinical cases (8.5 ± 2.1 days), later transitioning to a slight lagging indicator in December/January

2020-21 (-2.0 ± 1.4 days). A viral hotspot from within a larger catchment area was detected, prompting targeted interventions to successfully mitigate community spread; reinforcing the importance of sample collection within the sewer infrastructure. I conclude that by working in tandem with traditional approaches, WBE can enlighten a comprehensive understanding of population health, with methods and strategies implemented in this work recommended for future expansion to produce timely, actionable data in support of public health.

DEDICATION

This dissertation is dedicated to every door I've ever encountered, opened or closed, that has led me to this accomplishment.

ACKNOWLEDGMENTS

It has been an absolute privilege to learn from, and work alongside, the many creative and innovative minds here at Arizona State University, and to have experienced the numerous opportunities presented to me throughout my education and training. First and foremost, I would like to thank my co-advisor, Dr. Rolf Halden, for providing the opportunity to apply my background in nutrition towards an engineering degree; encouraging me to "think like an engineer." His unique ability to transcend through multiple disciplines and forecast an innovative solution is inspiring and humbling. I would also like to thank my co-advisor, Dr. Rosa Krajmalnik-Brown, for providing the opportunity to expand my laboratory skillset and knowledge-base into the fields of microbiology and molecular biology in order to explore new pathways in my research. Dr. Krajmalnik-Brown serves as a strong role model for women in STEM-related fields that I seek to emulate in my own career. I would also like to thank my committee members, Dr. Corrie Whisner, Dr. Arvind Varsani, and Dr. Otakuye Conroy-Ben who have each uniquely contributed to challenging me to think about my research in diverse aspects, and whose mentorship and support has been truly invaluable. I'd also like to thank Dr. Kaushal Rege for his guidance and support in the Biological Design program as a Graduate and Alumni Affairs Committee representative.

I would like to take this opportunity to acknowledge the many colleagues, classmates, and collaborators, who have each enhanced and shaped my experiences here at Arizona State University, in no particular order: Erin Driver, Temitope Oluwasegun Cephas Faleye, Hanah Goetz, Joshua Steele, Phil Choi, Michelle Young, Blake Dircks, Kerry Hamilton, Srivatsan Mohana Rangan, LaRinda Holland, Simona Kraberger,

Rafaela Salgado Fontenele, Rosa Inchausti, Wydale Holmes, Stephanie Deitrick, Matthew Scotch, Sayalee Joshi, Melanie Newell, Marcia Spurlock, Carole Flores, and Sarah Arrowsmith. I would also like to thank all of the undergraduate student mentees whom I have had the great honor of teaching and mentoring throughout my time, in no particular order: Jillian Wright, Sonja Savic, Bridger Johnston, Allison Binsfeld, Akhil Mahant, Caitlyn Fain, Jake Zevitz, Aly Carlson, Kathleen Click, Payton Watkins, Kaxandra Nessi, Nicole Campos, Mira Marwah, and Dustin Pollard. You all have taught me more than you know and solidified my calling for academia.

Chapters 4 and 5 would not have been possible without our dedicated collaborators with the City of Tempe, particularly in the Water Utilities Division, namely, Richard Dalton and Chris Garcia, along with all other participating field technicians who assisted with sample collection and handoff. Understanding flow dynamics and other pertinent and relevant information was made possible by Philip Brown, to whom I also extend my gratitude. I would also like to thank the Genomics Core at The Biodesign Institute, specifically Juan Maldonado Ortiz, for their assistance in sequencing samples as part of the analysis performed in Chapter 4.

This work has been funded in part by the following, to whom I extend my thanks: National Science Foundation, Virginia G. Piper Charitable Trust, The Flinn Foundation, The Catena Foundation, The J.M. Kaplan Fund through the J.M.K Innovation Prize, Jacobs Engineering, The School of Engineering of Matter, Transport, and Energy, through the University Graduate Fellowship, and The National Institutes of Health.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

Conventional strategies for assessing human nutritional intake, exposure routes to potentially harmful chemicals, and chronic and infectious disease are understood to be time-consuming, expensive, and resource-exhaustive [1, 2]. Diseases and conditions indicated to be influenced by environmental factors and/or human behavior and cause great healthcare and financial burden, such as obesity, type 2 diabetes, and malnutrition could be considered preventable, or long-term adverse effects minimized, if detected early in development [3]. Thus, it is widely encouraged to participate in preventative or early-detection public health programs, such as annual visits to see a doctor, however, these strategies do not always take consider certain factors such as cultural competence or socioeconomic status, and can also become quite labor-intensive. Further, despite efforts to make connections between human activity and health outcomes, both at the individualized and population-level, disease incidence and healthcare burden continue to persist and grow [2, 4]. Thus, this indicates a need for more innovative strategies that operate at scale, are contextual across regions, and offers a translational scientific advantage in order to achieve a pace comparable with disease prevalence. Wastewater-based epidemiology (WBE), an emerging scientific discipline that leverages community wastewater to gain insight into population-level health, has expanded widely in recent years and proposed as a beneficial methodology to be exercised in tandem with current methods across a multitude of public health applications [5-8]. Detailed discussion of these applications and the role of WBE is provided within.

1.1 Current methods for nutrition and disease assessment

The importance of conducting clinical nutrition research in order to draw causal links between dietary behavior and human health outcomes has become more widely realized and practiced as prevalence of nutrition-related chronic disease continues to exacerbate [9]. The gold standard for conducting a clinical trial, whether for nutrition or chronic disease, is generally an intervention-based investigation, such as a randomized controlled trial (RCT), which can allow for strong causal relationships to be identified between an intervention and the observed outcome through randomized selection of participant groups [10, 11]. Although these links are highly useful for clinical research, especially when applied to chronic disease, challenges such as extended study duration, large cohorts or multiple study sites, and strict exclusion criteria should all be carefully considered [11]. All of these factors, in addition to personnel time and labor, can contribute to costs, which can range anywhere between $4-20 million U.S. dollars per study conducted [12]. Predominant methods for assessing nutrient intake at population-level are survey-based, such as a '24-hour food recall or food frequency questionnaires' (**Table 1**) [1, 13, 14]. While these tools can potentially offer detailed insights into dietary patterns and behavior, challenges such as recall bias, participant fatigue, and inability to achieve comprehensive cultural competence may exist [4, 15, 16].

In addition to clinical trials that can aid our understanding of individualized human metabolism and response to interventions, there are several nationwide surveillance systems that exist as broad-scale tools for understanding population-level disease and behavior through both survey-based and selective biomonitoring efforts [3]. The Centers for Disease Control and Prevention (CDC)) oversees several disease surveillance systems

and programs, including both chronic and infectious diseases, in order to grasp the extent of human behavior and connection to health risk, monitor the progress of preventative efforts, such as education, and assist public health professionals and policy makers in downstream decision-making [3, 17]. However, the limitations experienced by these large-scale surveillance methods, are similar to those at the clinical-level, such as susceptibility to over- or underreporting, lack of study sample representation, and lag times for data turnaround [18]. Each of these systems are also individually managed, which could generate gaps across data sets. Thus, these challenges provide the opportunity to seek alternative and innovative strategies to continue to bridge gaps in data acquisition and reporting that is inclusive, cost-effective, and complementary to existing efforts.

Table 1. Definitions, advantages, and limitations for commonly-used conventional methods of dietary assessment in nutritional practice.

Tool	Purpose	Advantages	Limitations	Ref
24-Hour Recall	Report foods and beverages consumed in previous 24 hours	Standardized; least biased of all survey methods; detailed	Under-/over-estimation; misrepresentation of portions; often requires administration by trained professional	[19, 20]
Screener	Record basic information foods and beverages	Cost-effective; rapid, can support other instruments	Systematic error, often not culturally-relevant	[19]
Food Diary (a.k.a. Food Record)	Records all food and beverages consumed over one or more days	Long-term dietary intake; can be used via mobile apps	Thorough description required; under-/over-estimation	[19-21]
Food Frequency Questionnaire (FFQ)	Reports usual frequency of a food item for a specified period of time	Can reveal frequency patterns of foods/food groups	Systematic error; requires calibration; often excludes culturally-relevant foods	[19, 22, 23]

1.2 Wastewater-based epidemiology for population-level health assessment

In this thesis, a governing hypothesis is being tested that wastewater-based epidemiology (WBE) could serve as a valuable solution to the above-mentioned challenges, as it potentially may offer unique insights into population-level behavior, activity, and exposure by analyzing human excreted biomarkers in untreated, composited municipal wastewater [6, 24-26]. Historically, this methodology has been repeatedly implemented at an international level by employing established analytical techniques to understand population-level illicit drug use, typically by collecting samples at a wastewater treatment plant (WWTP) [27, 28]. Due to the success of this endeavor, studies have proposed to branch the field into other applications of human health, with few groups who have actively investigated the feasibility of determining specific indicators of disease and nutritional status in municipal wastewater [29-31]. While this thesis work was conducted, the coronavirus disease 2019 (COVID-19) global pandemic, caused by the virus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), caused a remarkable interest in, and appreciation for, the benefits of applying WBE as a complementary monitoring method for population-level health assessments beyond drug use. Consequently, WBE has now been applied in many new locations across the globe, and has experienced a rapid shift from almost exclusively focusing on chemical biomarkers to targeting biological signatures (i.e., RNA) in order to provide a comprehensive analysis of the presence of SARS-CoV-2 in communities [32-36].

This extended successful application of WBE indicates a much broader future use of wastewater analysis to inform our understanding of population health as a function of demographics, socioeconomic status, or geographic location. Due to the silent nature of

many chronic diseases in terms of noticeable symptoms, it may take the affected individual, or physician, if applicable, too long to recognize it early enough in order to prevent irreversible damage. In WBE, longitudinal and inclusive measurements of entire communities can work to establish baseline values and monitor ongoing trends, thus allowing for the early detection of poor human health outcomes at the initial sign of an increase beyond the established clinical threshold [18, 37, 38]. This use may promote a more targeted approach for public health strategies, enhancing preventative measures and allowing the opportunity to test the efficacy of interventions in near real-time; another unique quality and benefit to WBE.

1.2.1 Current gaps in wastewater-based epidemiology data

While there are reports of measuring dietary-relevant compounds in municipal wastewater at the city-level [39], there are currently no reports of a longitudinal and high-resolution (i.e., within sewer network) WBE study to monitor dietary behavior. Sample collection at the sewer catchment (i.e., neighborhood) level, may allow for deeper insights into consumption variability (daily, weekly, etc.) and could be beneficial for assisting current methods to capture dietary patterns and draw links to chronic disease. A unique attribute of food consumed by humans are the interactions with the human gut microbiome, yielding microbial products that are linked to certain aspects of human health and disease risk [40, 41]. As such, the human gut microbiome is increasingly gaining more attention within the scientific community, especially for precision nutrition and precision health purposes [42]. To date, no studies have explored utilizing WBE to understand human gut microbial interactions as it relates to population-level diet, or as a method to inform or provide further context to chemical measurements relevant to

microbial activity. Information learned from this investigation could unlock even more opportunities to broaden the field of WBE and inform targeted population health strategies and interventions.

Longitudinal assessments can allow for the establishment of measuring robust baseline values, and combined with neighborhood-level sample collection, it can provide a highly contextual assessment of the specific community served. Using this methodology, a threat, such as an emerging infectious disease or a rising indicator of poor human health, can be detected and subsequently mitigated in near real-time. Further, wastewater surveillance from within the community may allow for early warning potential of an emerging threat, whether a new infectious disease or rising indicator of poor human health, which could be used in tandem with conventional methods to strengthen the public health response. While prior studies have explored this early warning capability, few, if any, have explored public health outcomes observable by WBE as a function of demographics, socioeconomic status, and geographical location, which can be crucial to design, implement, and monitor the success of appropriate and relevant intervention strategies to properly reduce or eradicate the threat [35, 36, 43].

1.3 Dissertation goals and research strategy

Thus, the goal of this dissertation was to explore these broader applications of wastewater-based epidemiology (WBE), both theoretically and experimentally, and critically evaluate their potential as a complement to current methods for understanding the connection between population-level human behavior, exposure, and activity and related diseases.

In an extensive literature review, I examined the potential for performing WBE to understand population-level nutritional status; the ultimate goal being to introduce this concept to the field of WBE from the perspective of nutritional science. As mentioned, few studies had investigated this new application for WBE, thus it was crucial to explore the literature and obtain a list of potential biomarkers deemed feasible for subsequent method development and experimentation using liquid chromatography-tandem mass spectrometry (LC-MS/MS). This goal was achieved when a short list was created based on literature-reported values largely from clinical studies of detectable amounts in human excrement, predominately urine, that represented a wide range of food intake (meat, plants, whole-grains, etc.). Further, analysis of these biomarkers in terms of their human health and physiological relevance was explored by examining accompanied odds ratios to specific types of chronic diseases, such as type 2 diabetes. A second in-depth literature review was performed that explored exposure to estrogen-mimicking endocrine-disrupting chemicals exclusively through ingestion of various foods that either naturally contain these types of chemicals (phytoestrogens), are indirectly exposed by interactions with plasticizers in food packaging, or coated with trace amounts of pesticides. Reported estimated ingestion rates for some food sources were comparable to reports of body burdens of EEDs when examining causal links to breast cancer incidence. Interactions with the estrogen receptor (ER) and physiological disruptions to breast tissue as a function of low, chronic exposure to these compounds was also explored. The purpose here was to demonstrate the direct link between human dietary behavior and health outcomes.

The comprehensive information obtained from the previous two literature assessments was incorporated to inform subsequent experimentation in a longitudinal WBE study. Three phytoestrogens (genistein, daidzein, and enterolactone) that represent two major classes of human-consumed phytoestrogens (isoflavones and lignans) were monitored for two years in a small catchment within a larger city to understand population-level dietary behavior and consumption patterns informed by wastewater analysis. A novel microbial metabolite of daidzein (equol) was added later in the study to test parent-metabolite behavior and propose feasibility for integrating into future studies. Due to human phytoestrogen metabolism involving the human gut microbiome, as proof of concept, a microbial composition analysis on a subset of samples was performed using WBE, targeting the 16S rRNA bacterial gene. In order to test the ability to investigate potential human gut microbiome interactions, relative and semi-quantitative abundances of select microbial genera involved in phytoestrogen metabolism was measured and calculated. Results suggest this approach to WBE is feasible, however, further exploration is much needed as the human gut microbiome is but a small representation of the many microbiomes existing within community wastewater (stormwater, animals/pets, biofilms within sewage pipe network, etc.). Nonparametric statistical analyses were performed to determine significance of changes in consumption patterns.

Finally, the COVID-19 global pandemic during which this thesis work was completed highlighted many challenges and opportunities for improvement in monitoring for and management of infectious diseases at population-scale. High-frequency (3x/week), high-resolution (multiple collection sites) wastewater sample collection was implemented in a year-long study to assess changes in virus presence throughout a

southwestern U.S. city in an effort to support public health strategies and inform targeted interventions. Population health data informed by wastewater were reported weekly and posted to a publicly-facing, online dashboard in order to promote data transparency and enhance public health efforts. Statistical analyses were performed to test the early warning capability of WBE in comparison to newly reported clinical cases, and COVID-related hospitalizations and deaths reported at the zip code or county-level.

1.4 Hypotheses

Overarching research goal: Important determinants of human health, such as dietary behavior, and chemical and biological markers linked to acute and chronic diseases, are detectable in wastewater at the community level and, therefore, is feasible to harness a population-level health assessment to gauge the impact of public health interventions implemented.

Specific hypotheses:

(*i*) biomarkers indicative of dietary behavior are detectable in municipal wastewater and associated with nutrition-related chronic disease defined by an odds ratio (OR) that indicates positive (OR>1), negative (OR<1), or no (OR=1) association;

(*ii*) ingestion rates of foods that contain EEDs at increased amounts are associated with disease burden in terms of breast cancer incidence than those consumed in lower amounts as reported by literature values;

(*iii*) per capita consumption (mg d^{-1} per capita) of isoflavones genistein and daidzein and production of enterolignan, enterolactone, measured in wastewater collected

from within a small sewer catchment exhibit statistically significant seasonal trends (α=0.05);

(i*v*) per capita production (mg d^{-1} per capita) of human gut microbial metabolite, equol, measured in wastewater collected from within a small sewer catchment are positively correlated ($r > 0.50$) with daidzein (md^{-1} per capita);

(*v*) measurable signals of the E gene of SARS-CoV-2 detectable in wastewater precede newly reported clinical cases, COVID-related hospitalizations, and deaths, thereby providing an early-warning capacity useful to inform the public health response.

1.5 Specific aims

Specific aims of this dissertation were to:

(*i*) Determine feasibility of measuring nutrition- and disease-related biomarkers in a wastewater matrix through extensive literature analysis and examine possible interpretations of those measurements in terms of population health;

(*ii*) Measure indicators of a plant-based diet (i.e., phytoestrogens) in a small residential population within a southwestern U.S. city over time to discern a relationship between wastewater-derived measurements and estimated reports of per capita consumption;

(*iii*) Monitor SARS-CoV-2, the virus that causes COVID-19, in wastewater at neighborhood-level within a southwestern U.S. city to assist current methods for public health surveillance, understand viral presence and transmission, and compare trends with reported data for new clinical cases and COVID-related hospitalizations and deaths;

(iv) Provide a foundation to inform future work for utilizing wastewater-based epidemiology to develop population-level health assessments; used in tandem with current epidemiological methods.

CHAPTER 2

THEORETICAL EVALUATION OF USING WASTEWATER-BASED EPIDEMIOLOGY TO ASSESS THE NUTRITIONAL STATUS OF HUMAN POPULATIONS

This chapter was published in an altered format in *Current Opinion in Environmental Science and Health [44]*

Introduction

This dissertation encompasses the work of individual studies that serve to critically examine the employment of wastewater-based epidemiology as a viable and complementary tool for population-level health assessments, particularly as it relates to nutrition and chronic and infectious disease.

In individualized human health studies, clinical trials are typically conducted to determine a relationship between a particular behavior, activity, or exposure and an outcome, especially when examining links to chronic disease. A randomized controlled trial is often used as it serves to reduce bias, isolate the intervention, and determine potential causality. While these types of studies provide in-depth insight into human disease incidence, study duration may last for several years, which can become quite costly. Additionally, results reported in the literature may not always reach the appropriate audience that stands to profit from the data collected. Further, clinicians who operate in private practice are encouraged to report incidence of new diseases, such as type 2 diabetes, however, this is not always enforced which commonly leads to underreporting.

National monitoring systems for chronic disease risk and infectious disease exist and are mandated predominantly through large government agencies, such as the Centers for Disease Control and Prevention (CDC). These systems can serve to fill in gaps where clinical studies may be limited, while also providing high-level insight into connections between particular behaviors and disease prevalence. Many of these programs operate using survey-based methods along with selective participants to undergo more in-depth analysis and, at times, biomonitoring. While useful, there are still instances of over- or underreporting due to the self-reporting nature of survey-based methods, and some types of survey instruments are not culturally competent, which could lead to underrepresentation.

In Chapter 2, these challenges for conducting human health assessments as it relates to nutrition are examined in-depth, with a proposed alternative to understanding population-level nutritional status: wastewater-based epidemiology. WBE was hypothesized to potentially provide the ability to conduct population-level assessments, while providing objective results from the contributing population by analyzing human excreted biomarkers indicative of behavior, activity, and exposure in a community wastewater sample. In Chapter 2, chemical signatures indicative of dietary behavior were identified through extensive literature analysis, including representation of plant, meat, and fiber. Their feasibility for detection in a wastewater sample by liquid chromatography-tandem mass spectrometry (LC-MS/MS) was assessed based on reports of urine and/or fecal analysis at the individual level, and connection to chronic disease was determined based on literature-informed odds ratios. A list was compiled and served as a foundation for subsequent experimentation using WBE.

2.1 Current dietary assessment methods

The predominant method for observing the nutritional history and dietary trends of individuals and human populations is to conduct surveys in which respondents self-report on the type and quantity of food items consumed in a particular timeframe [2, 45]. For example, food frequency questionnaires (FFQ), 24-hour recall surveys and food diaries may be distributed to recipients for recounting of nutritional intake over a specified period of time (e.g., 1-7 days) [1, 2]. This approach has served to observe associations between symptoms experienced by humans upon consumption of food items either within that same day or over a longer time period [2]. While these modalities may be relatively inexpensive and simple to conduct, they are known to suffer from limitations associated with participant bias and inaccurate assumptions, such as subjective views of food quantities ingested or cooking practices employed [2]. Moreover, data obtained from one respondent at a time may or may not be reflective of behavior said person in the past, in the future or among other individuals with similar demographic characteristics [1, 2]. For this reason, engaging a large number of individuals in a given survey may reveal population-level information and would be beneficial for deeper analysis of understanding nutritional trends [6]. Furthermore, due to the sensitive nature of discussing diet, recalling information in this manner can also render surveyed individuals uncomfortable, requiring extra attention and potentially lengthy appointments [1]. Additionally, some physicians are not adequately trained in specialized nutritional counseling and facilitating behavior modification strategies, thereby resulting in an inefficient use of resources and the potential for acquisition of inaccurate data [1]. Use of anonymized data may preserve respondents' anonymity and

ultimately may aid in establishing linkages between dietary intake and outcomes in individual health and population health.

2.1.1 Dietary associations to adverse health outcomes in human populations

Chronic diseases, such as Type 2 Diabetes Mellitus, are progressively contributing to increased rates of morbidity and mortality globally, thereby exacerbating healthcare costs and losses in productivity and quality of life [46]. In 2017, this disease alone affected 425 million people worldwide, and it is projected to affect nearly 630 million people annually by 2045 [1]. This disease may be prevented or managed rather effectively by addressing known risk factors, such as adopting a more balanced diet and increasing the frequency and duration of physical activity [46]. Yet, it is difficult to implement these adjustments in clinical practice, due to biased information from self-reporting patients and differences in expectations among physicians and survey participants [1].

One health outcome with known linkages to nutrition is cancer. For example, several studies have indicated that consumption of red meat is positively linked to the onset of several cancers, including those of the colon (odds ratio, OR = 1.17; 95% confidence interval, CI: 1.08-1.26) and the rectum (OR= 1.22; 95% CI=1.11-1.33) (**Table 2**) [47, 48]. Breast cancer is the second leading cause of death from cancer in female populations in the United States [49]. Evidence suggests that soy intake as measured by the isoflavones genistein and daidzein (OR= 0.57; 95% CI=0.39-0.84), and that lignans, assessed as enterolactone (OR= 0.82; 95% CI=0.69-0.97), exhibit protective effects against malignant growth of breast tissue (**Table 2**) [49, 50]. Interestingly, other evidence suggests positive associations between malignant growth of breast tissue and dietary consumption of phytoestrogens [51]. The isoflavones genistein and daidzein are

understood to act on the estrogen receptor as estrogen mimics at the estrogen receptor (ER), and are termed estrogen-mimicking endocrine disruptors or EEDs; however, further study is needed to establish the causality between associations observed between dietary intake of these food items and incidence of breast cancer in women [51-54]. Reports of contrasting associations call for further examination of nutritional choices and prevalence of chronic disease [51].

Table 2. Odds ratios (OR) and 95% confidence intervals observing diet and chronic disease.

Food	Human Metabolite	Approximate Consumption (g/day)	OR (95% CI)	Disease	Association	Source
Meat	1-Methylhistidine	60-89	1.19 (1.02-1.38)	Colon Cancer	Direct	[47]
		60-89	1.25 (1.04-1.51)	Rectal Cancer	Direct	[47]
		>85	1.97 (1.10-3.55)	Type 2 Diabetes	Direct	[48]
Isoflavones	Genistein, daidzein, equol	0.008–0.016	0.57 (0.39-0.84)	Breast Cancer	Inverse	[49]
Lignans	Enterolactone	0.16	0.82 (0.69-0.97)	Breast Cancer	Inverse	[50]

2.2 Wastewater-based epidemiology as a complementary method for assessing dietary intake

Wastewater-based epidemiology (WBE) is an inexpensive tool that relies on the detection in community wastewater of human metabolites (e.g., biomarkers) excreted by local residents into the municipal sewerage system to estimate the per capita consumption of chemicals and possibly in the future, of the human diet [6]. Customarily, biomarker

concentrations measured in composited community wastewater are multiplied with daily flow rates to arrive at an estimated mass load of chemicals per unit time [55]. These mass estimates can then be divided by the size of the population served by the wastewater treatment plant to obtain estimates of average chemical intake per 1,000 persons [6, 56]. Wastewater contains a wide variety of biological and chemical markers excreted in urine and feces, thereby presenting a diagnostic matrix that may provide insights into both population health status and human activity [6, 56]. In order to quantify estimates of a metabolite of interest, the samples typically are collected using automatic samplers over a specified period of time, e.g., 24 hours, and are then analyzed using liquid chromatography-tandem mass spectrometry [6, 55]. This approach allows for the near real-time observation of very large populations of up to a million people or more; with processing and analyzing within a short-period of time (e.g., 48 hours) and reported back to parties of interest [57]. Thus, population-scale health information may be obtained, analyzed, and disseminated in a cost-effective and timely manner [6, 55].

The WBE approach has been successfully applied to investigate trends in substance use (e.g., illicit drugs), lifestyle risk factors (e.g., consumption of alcohol, nicotine and caffeine), exposure to environmental toxicants of concern (e.g., bisphenol-A (BPA), pesticides, and parabens), and the assessment of stress levels through hormonal fluctuations [6, 55, 57-59]. It has been hypothesized that the WBE approach also could be applied to inform on nutritional status; however, further studies demonstrating these biomarkers' stability in wastewater and urinary excretion ratios are needed [57]. Use of WBE may enable one to track metabolites as exact indicators of consumption of food stuffs known or considered to represent a risk factor for diseases, including the burdening

chronic diseases associated with high healthcare costs and significant rises in morbidity and mortality rates (**Figure 1**) [6, 8, 56].

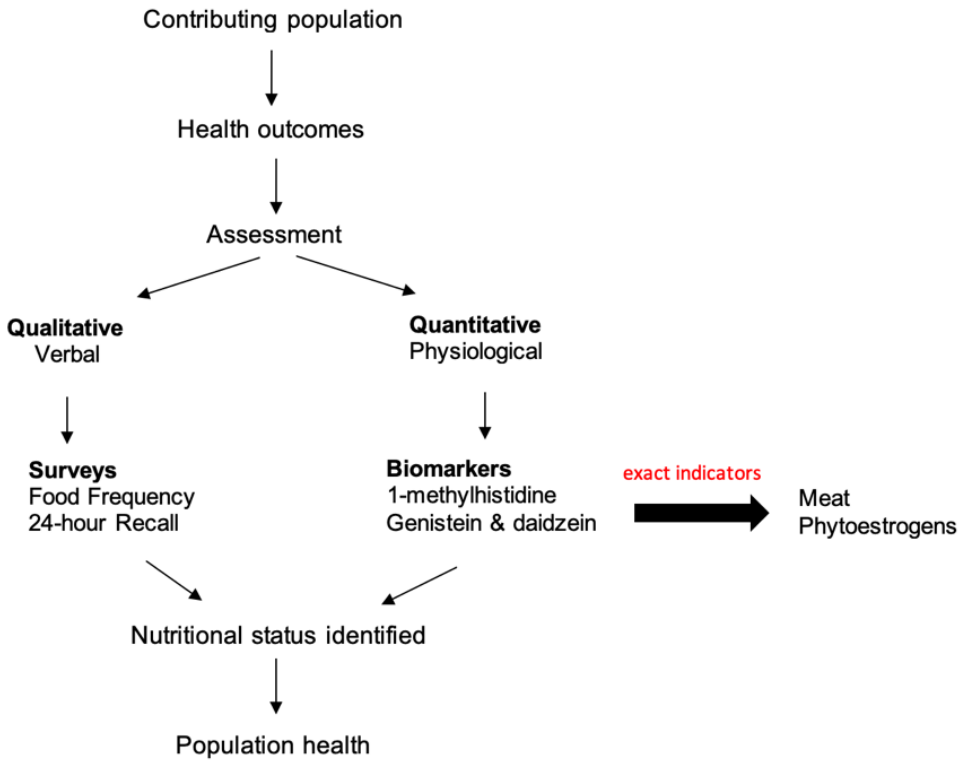

Figure 1. Schematic of envisioned use of wastewater-based epidemiology as a comprehensive tool for assessing nutritional status of a population.

2.2.1 Target analytes

Current literature documents that the consumption of certain food items and food groups is positively associated with the incidence rates of chronic diseases (**Table 2**) [47-50]. For example, intake of red meat was found to be correlated with an increase in the incidence of certain cancers, including of the colon and rectum, as well as with the onset of type 2 diabetes [1, 46, 60]. Anserine is a compound found exclusively in the tissue of animals consumed regularly by humans (e.g., chicken, turkey, beef) [61]. Upon ingestion by humans, the enzyme carnosinase can break down anserine into two major metabolites,

1-methylhistidine and β-alanine[61]. β-Alanine is a naturally occurring amino acid that plays a role in many processes of human physiology [61]. Since 1-methylhistidine is not utilized further in any metabolic pathway, it is readily excreted in urine; therefore, 1-methylhistidine would make for a potentially valuable biomarker that is trackable by WBE and that potentially could indicate the average consumption in a surveyed population of all meats, including red meat, poultry, pork, etc. (**Table 3**) [8, 62, 63].

Phytoestrogens are plant-derived compounds belonging to different chemical classifications; isoflavones, lignans and coumestans represent three major families of phytoestrogens present in the human diet [51, 53, 54]. Collectively, these three subgroups exhibit estrogenic as well as anti-estrogenic properties to varying degrees by acting at the estrogen receptor as either estrogen inhibitors or estrogen mimics [51, 53, 54]. Genistein and daidzein account for approximately 95% of the isoflavone content of soy foods, identifying them as potentially valuable analytes for studying the consumption of popular food products such as tofu, miso and soymilk (Table 3) [51, 54]. Equol, a metabolite of daidzein catalyzed by specialized microbes of the human gut microbiome, is projected to be produced in approximately 30-40% of the population [51, 53, 54]. Equol is considered to have greater estrogenic potency than genistein or daidzein, possibly due to its structural similarity to endogenous estrogen; thereby offering either chemo-protective mechanisms or carcinogenic effects to the host [53].

Unlike isoflavones, lignans are more ubiquitous throughout various non-soy food sources such as seeds, whole-grains, fruits and vegetables (e.g., berries, broccoli, and kale) [64]. Enterolactone, a microbial product of the intestinal digestion of matairesinol and secoisolariciresinol (two dietary lignans), amongst others, could serve as a biomarker

for consumption of these foods, as it presents at high levels in human urine upon habitual consumption of foods containing lignans [64].

Other compounds of interest that could provide more quantitative data on dietary consumption patterns include alkyl resorcinols (AR) such as 3,5-dihydroxybenzoic acid (DHBA) and 3-(3,5-dihydroxyphenyl-propanoic acid (DHPPA) for whole-grain consumption, and allyl isothiocyanates (AITC) for cruciferous vegetable consumption (Table 3) [45, 65]. Examining by WBE the excretion patterns of the metabolites DHBA and DHPPA for whole-grain consumption would provide complementary qualitative measurements alongside traditional FFQs [45]. Whole-grains provide many nutritional benefits including contributing to overall daily intake of fiber, which has been shown to improve colon function and cardiovascular health [45]. Consumption of cruciferous vegetables at notable quantities is part of a healthy, well-balanced diet, as these offer a wide variety of essential vitamins and minerals, thereby decreasing risks of developing chronic diseases [66, 67]. On many occasions, patients may unintentionally report higher intake of vegetables than what other modalities may indicate [67, 68]. It has been stated that a lack of knowledge of proper serving sizes of nutritionally dense vegetables represents a major barrier to living a healthy lifestyle [66]. Monitoring AITCs as biomarkers that are indicative of these nutritionally dense cruciferous vegetables would complement the information obtained solely by qualitative surveys.

Table 3. Identified human biomarkers proposed for WBE indicative of various types of foods.

Parent Compound	Human metabolite	Indicator	Source
Anserine	1-MH[1]	Total meat intake	[8, 62, 63]
Daidzein	Daidzein (parent)	Isoflavone intake	[51, 53, 54]
Genistein	Genistein (parent)	Isoflavone intake	[51, 53, 54]
Matairesinol & Secoisolariciresinol	Enterolactone	Lignan intake	[64]
Daidzein	Equol	Reflect the percentage of a population who can produce equol	[53]
Alkylresorcinols	DHBA[2], DHPPA[3]	Whole grain wheat and rye intake	[45]
Allyl isothiocyanates (AITC)	NAC-AITC[4]	Cruciferous vegetable intake	[65]

(1) 1-MH: 1-Methylhistidine; (2) DHBA: 3,5-Dihydroxybenzoic acid; (3) DHPPA: 3-3,5-Dihydroxyphenyl)-propanoic acid; (4) NAC-AITC: N-acetyl-S-(N-allylthiocarbamoyl) cysteine

2.2.2 Potential limitations

Wastewater-based epidemiology is a tool that may aid in explaining observed differences between self-reported dietary intake and actual food consumption [6, 56]. Data obtained by WBE are predominantly quantitative, which could serve to complement the qualitative outcomes of food diaries and self-reported surveys [1]. Although the use of WBE may provide a more comprehensive view of the nutritional status of a given population, this approach also has limitations, including a still limited spectrum of potentially useful biomarkers (**Table 3**) and the need to validate these prior to use. Food intake markers and metabolites may be subject to in-sewer degradation post excretion, such as microbial activity that could break down the parent compounds into the identified human metabolites, which would potentially introduce a certain level of bias needed to be

taken into consideration [57]. Furthermore, the parent compounds specifically for the intake of the isoflavones genistein and daidzein (**Table 3**) are measured and accepted as compounds of human consumption. It is preferable when utilizing WBE to detect the human urinary metabolized compounds in order to prevent overestimation of consumption rates; therefore, this poses as a quantification limitation for these analytes [6]. By only investigating population-level results, this approach smooths out the maxima and minima of individual consumption, allowing only an assessment of the average intake, which may not be characteristic for any one of the individuals captured by this approach [55]. While representing a limitation on the one hand, this feature also is appealing from an economic standpoint and from the vantage point of preserving anonymity of those surveyed [56]. Estimated per capita consumption rates of target analytes are based on population estimates whose accuracy may differ widely, depending on the approach taken (e.g., census data for population size in the sewershed versus mass of caffeine as a proxy for the number of individuals represented in a sample) [6, 69].

It is also important to note that while some metabolites of interest are unique to humans, there is also the possibility of capturing metabolites from animal excretion as well as runoff from nearby restaurants or related facilities; dietary compounds being especially vulnerable. Due to these limitations, analyzing sewage sludge as a diagnostic matrix for those dietary markers that are hydrophobic and persist to a sufficient degree to get sequestered in sludge could serve as another potential method for quantifying these compounds and assessing the human health impact [70]. While WBE has become an accepted tool for studying the consumption of licit and illicit drugs, e.g., in major European cities, its use for assessing dietary preferences and associated human health

status and disease prevalence have yet to be explored [71]. The present theoretical analysis points to both the feasibility and potential benefit of such future investigations.

2.3 Conclusions

Monitoring wastewater-borne human metabolites indicative of dietary trends within a population is projected to be both feasible and practical, showing promise to allow for a better assessment when used in conjunction with traditional survey tools. Conventional methods provide qualitative data whereas WBE may add aggregated quantitative information on average consumption. Thus, enahncing the confidence of dietitians when assessing dietary habits known or presumed to affect human health status. Chronic diseases such as Type 2 Diabetes and various types of cancers place a great burden on the healthcare system and on the quality of life of individuals, translating to financial, emotional, and mental stress in human populations. Employing WBE as an inexpensive, non-invasive and time-efficient tool that preserves the anonymity of those surveyed promises to provide valuable information regarding the population observed. Decreased ambiguity will allow for elucidating linkages between nutrition-related diseases and their dietary sources in addition to prior work focusing on consumption of illicit drugs and exposure to contaminants of emerging concern [56, 72].

CHAPTER 3

BREAST CANCER AND DIETARY INTAKE OF ENDOCRINE DISRUPTING CHEMICALS: A REVIEW OF RECENT LITERATURE

This chapter was published in an altered format in *Current Pathobiology Reports [73]*

Introduction

Assessing trends in human excreted biomarkers indicative of dietary behavior, which can also serve as a secondary indicator of links to, or risk factors of, chronic diseases such as type 2 diabetes, have only just begun to be explored using WBE. At an individual level, biomarkers can be measured by collecting a single biological specimen, (e.g., urine, feces, etc.) and tested against established reference ranges to determine if a disease or condition may be present. This type of intensive and potentially invasive procedure is not performed frequently (i.e., annually), and may deter individuals from seeking care. For understanding nutrition, survey-based methods are validated and effective when attempting to gain insight into dietary behavior both at the individual- and population-level, however, the limitations that exist (recall bias, under- or overreporting) can create gaps in data acquisition. Wastewater-based epidemiology is positioned as a feasible alternative to understand, at population-level, nutritional status and dietary patterns of a community. The proposed biomarkers in the previous chapter create a foundation for future investigation for monitoring at any scale (neighborhood-level, treatment plant, etc.) in order to understand how this tool can be used as a complementary source of information for public health strategies.

In Chapter 3, I performed an in-depth literature analysis to understand exposure to potentially harmful chemicals through consuming foods either naturally containing or

contaminated with endocrine-disrupting chemicals that specifically mimic estrogen, termed, estrogen-mimicking endocrine disruptors (EEDs). Evidence suggests a link exists between exposure to EEDs and breast cancer incidence, however, investigation of the ingestion exposure route through food had yet to be explored in detail. I explored the recent literature (≤5 years) to identify exposures to EEDs through food consumption. Several types of EEDs were identified, including components of plasticizers in food packaging, naturally-occurring EEDs, and pesticide residues, and subsequently cross-examined to determine potential mechanistic behaviors and feasibility of contributing to breast cancer incidence. Reported body burdens from studies investigating levels of exposure of EEDs to physiological changes either in human or animal studies were explored.

3.1 Estrogen-mimicking endocrine disrupting chemicals and types of breast cancer

Endocrine disrupting chemicals (EDCs) are ubiquitous substances found in our food, the environment, and in purchased products that can interfere with hormone signaling, metabolism and action, leading to a disruption in hormone homoeostasis and potentially a higher incidence of breast cancer [74, 75]. Estrogen-mimicking endocrine disruptors (EEDs) found in food and non-food items are a subgroup of EDCs that bind to estrogen receptors, despite their low concentrations [76]. Breast cancer is of particular interest when studying EEDs due to breast tissue development being heavily reliant upon the physiological fluctuations of endogenous estrogens, namely estradiol [77]. There are three major types of breast cancer: hormone-responsive (HR+), human epidermal growth factor receptor 2 (HER2+), and triple-negative. Regardless of the type, it is estimated that

dietary factors may contribute up to 30-35% of incidence [51]. Malignant growth classified as HR+ accounts for approximately 80% of all breast cancer cases, in which the cancer cells grow in response to estrogen levels [78]. Thus, this type is the most concerning in regards to studying exogenous hormone mimics interfering with breast tissues' typical mode of development, posing a high risk for irregular cell growth patterns and cancer cell proliferation. Specific EEDs implicated with increased cancer risk include Bisphenol A (BPA) [79] several pesticides (notably, dichlorodiphenyltrichloroethane (DDT) and atrazine [80, 81] and phytoestrogens (isoflavones genistein and daidzein) [51, 53] all discussed in the following.

3.1.1 Identity and potency of EEDs in the human diet

Bisphenol A is a synthetic compound widely used in many dietary consumer products including plastic food packaging and canned foods due to its ability to protect against corrosion, as well as food items such as cereals, baking powder, and yeast [79] (**Table 4**). BPA has been discovered to migrate into food directly from food packaging as a free, unconjugated monomer [82]. Factors such as agitation from heat, such as in the microwave, or adding hot water to pre-packaged, dehydrated foods (e.g., soups), can encourage chemical migration from the packaging into the food [79, 83]. Due to this characteristic, the European Union set a specific migration limit (SML) to 0.6 mg/kg food for manufacturers to adhere to when producing products with BPA to ensure a tolerable limit is not exceeded [82]. As set by the US EPA, the lowest observable adverse effect level (LOAEL) of BPA in humans is set to 50 mg/kg adipose tissue/day with a NOAEL (no observable adverse effect level) set to 5 mg/kg adipose tissue/day [84]. It is estimated that exposure to BPA via dietary ingestion in adults occurs at a rate of 0.4 - 4.2 μg/kg-

bw/day [79]. Studies conducted to test the concentrations of BPA in 175 different brands and types of canned foods in the US showed a BPA detection frequency of 91%, with levels ranging from 0.2-730 µg/kg dry weight of food. Results were highly variable among food types (e.g., tuna fish, green beans); however, there appeared to be a pronounced difference between brands (e.g., green beans exhibited a 30-fold difference (22-730 ng/g) [79].

Table 4. Dietary exposure of EEDs and estimated ingestion amounts.

EDC	Dietary Exposure	Average Ingestion (mg d^{-1})	Reference Dose (mg kg^{-1}-bw d^{-1})	Contributed through diet (%)	Source
Bisphenol A (BPA)	Canned foods, yeast, baking powder, cheese, breads, cereals	4.0×10^{-4} – 4.2×10^{-3}	5	90%	[79, 84]
Parabens	Cereals, milk, fish, seafood, beans, fruits	5×10^{-8} – 3.6×10^{-3}	10	unknown	[83, 85]
Isoflavones (Genistein and daidzein)	Soybeans, cereals, soy-based formula, imitation dairy products, meat substitutes	1 – 3	None	100%	[44, 53]
Pesticides (DDT & atrazine)	Fruits, vegetables, water	DDT: 3.0×10^{-5} Atrazine: 3.3×10^{-5} – 1.2×10^{-5}	DDT: 5.0×10^{-4} Atrazine: 3.5×10^{-2}	unknown	[86] [80] [87]
Phthalates	Milk, meat, fish, seafood, eggs, poultry	6.6×10^{-4} – 1.6×10^{-3}	2.0×10^{-2}	unknown	[81, 83] [85]

Among the many pesticides studied, DDT and atrazine are two that have been shown to cause estrogenic dysfunction through various mechanisms. Both are considered EEDs and displayed a positive correlation with breast cancer development [80]. Whereas DDT has been banned in the US since 1972, the parental compound and its major

transformation products (DDD, DDE) still remain and persist in the environment, with continued use in African countries for managing malaria [88]. In the US, the Agency for Toxic Substances and Disease Registry (ATSDR) states that dietary exposure to DDT today could result from ingestion of contaminated drinking water, fish and seafood containing small amounts in their tissues, as well as consumption of fresh fruits and vegetables grown on contaminated soil [87]. For DDT, the average daily intake (ADI) was estimated in the early 2000's to be 0.03 μg/kg/day with a RfD of 0.5 μg/kg/day [87]. It is assumed these levels will continue to decrease as time progresses. Atrazine is still in use today as an herbicide in some countries and notably the US, with maximum levels for food residues tolerated within the range of 0.02-15 ppm and maximum levels allowable in drinking water of 3.0 μg/L [87]. The US reference dose for atrazine is 35 μg/kg/day [89]. Typically, dietary exposure is attributed to drinking contaminated water from personal water wells or ingestion of residues on fresh fruits and vegetables [87]. Human data for ingestion rates is sparse; however, the National Institutes of Health Toxicology Data Network (NIHTDN) in a study from 1984-1987 reported an intake range of 0.033 – 0.0123 μg/kg/day [90].

Phytoestrogens differ from the aforementioned chemicals since they are naturally derived from plant-based foods. While it has been previously thought that phytoestrogens only promote the reversal of cancer cell growth and prevent cell proliferation, newer studies now suggest that chronic uptake of low amounts of soy, particularly soy isoflavones contained therein, could lead to disruption in estrogen levels during breast tissue development. This could lead to mutations later in adulthood from exposure during specific windows of susceptibility, such as during pre-pubertal stages or infancy from

soy-based formula intake [51]. The isoflavones genistein and daidzein, diphenolic non-steroidal compounds, are one class of phytoestrogens commonly found in soybeans, beans, and lentils as well as in products marketed as substitutes for meat and dairy, such as tofu, tempeh and soy milk [91] (**Table 4**). Genistein accounts for approximately 50% of isoflavone content in soybeans, daidzein about 40%, for a collective total accounting for up to 90% of isoflavone exposure and burden [92]. Isoflavones are known to be present in food naturally as inactive glycoside conjugates, typically with a carbohydrate such as glucose, and to become unconjugated into the bioactive form, aglycone, after interaction with the saliva which hydrolyzes the sugars and interacts with intestinal microflora [91, 92].

Increased isoflavone content in the human diet in the US could be a result of popular dietary trends, making it particularly interesting when examining intake of other regions of the world where consumption of isoflavones has been historically higher in comparison, however, the dietary sources are slightly different. For example, in countries such as Asia, isoflavone intake can range anywhere from 15,000-50,000 μg/person/day mostly from fermented soy products, whereas in the US, isoflavone consumption is about 1,000-3,000 μg/person/day [53, 92]. Each gram of soy protein in soybeans is estimated to contribute about 3,500 μg of isoflavones, therefore, 250 mL soymilk may equate to approximately 25,000 μg of isoflavone exposure [92]. In a study conducted to determine phytoestrogen content of 115 various animal-derived food products, an average of approximately 20 μg/100g wet weight comprised of isoflavones (6 μg/100g), lignans (6 μg/100g), equol (3 μg/100g) and enterolignans (6 μg/100g), were detected [93]. These

results indicate the possibility that human ingestion of phytoestrogens is greater than previously assumed, prompting further investigation.

Phthalates are also a major component of plastic food packaging typically used as softeners or plasticizers [94]. These chemicals are naturally lipophilic and not covalently bound to plastic packaging when present as additives, rendering them susceptible for easy migration into food [83]. Foods such as milk, meat, seafood, eggs and poultry (**Table 4**) all contribute to an overall ingestion exposure rate ranging from 0.66-1.61 μg/kg-bw/day [83]. These observed levels are much lower than the RfD via ingestion (20 μg/kg-bw/day) [95]. Exposure to parabens via ingestion is more difficult to delineate. Parabens have been used as preservatives in processed foods, such as dried meats, and are also naturally found in fresh berries [83] (**Table 4**). Exposure via ingestion of foodstuff has been quantified to an approximate range of 50 – 3,600 μg/g-bw/day, with consumption of fatty fish contributing to the larger amounts [83]. There is no RfD listed on the EPA regarding parabens, however the European Food Safety Authority (EFSA) has stated that an acceptable daily intake (ADI) of parabens is 10,000 μg/kg-bw/day [85] rendering the overall consumption generally safe.

3.1.2 Evidence linking dietary EEDs to breast cancer

In a study that tested the urine of 75 male and female volunteers who consumed one serving of canned soup per day during a five-day period, there was a 1200% increase in urinary BPA levels compared to concentrations after eating fresh food the previous five days. There was a 66% decrease observed after following three days post-canned food consumption [79]. This result suggests that there is still a large amount of BPA traveling throughout the body even three days after ingestion due to prolonged excretion.

BPA has demonstrated a LOAEL in rat studies of 50 μg/kg adipose tissue/day, leading to irreversible effects on the reproductive system and reduced body weight (**Table 5**) [84]. Earlier exposure in life corresponds to higher risk of breast cancer in adulthood due to breast tissue being particularly sensitive to mutations during puberty [81]. Multiple studies demonstrated that fetal exposure to low doses of BPA can cause proliferative effects and an increase in estrogen sensitivity [74]. These effects can decrease apoptosis and mutate mammary gland tissue. Rodents exposed to low doses of BPA during gestation show morphed mammary gland development in utero, leading to increased sensitivity to estrogen and progesterone in adulthood [74].

Table 5. Estimated body burden of estrogen mimicking endocrine disruptors acquired primarily by dietary exposure.

EED	Body Burden (mg/kg/day)	Study Subject	Effect	Source
DDT	0.2	Rats	Liver lesions, hepatocellular hypertrophy, peripheral fat storage (females)	[95]
Atrazine	25	Rats	Decreased weight gain, systemic toxicity, reproductive toxicity	[95]
BPA	0.05	Rats	Irreversible reproduction effects	[84]
Isoflavones	0.3	Humans	Carcinogenic activity of mammary gland	[53]
Parabens	1,000	Humans	Irritation to GI tract	[90]
Phthalates	19	Rats, Guinea Pigs	Increased kidney and liver weight	[95]

DDT and atrazine exposures have been historically linked to breast cancer via endocrine disrupting mechanisms [80], each having their own unique LOAEL (DDT: 200 μg/kg adipose tissue/day, atrazine: 25,000 μg/kg adipose tissue/day) observed in both humans and animals (Table 5) [95]. Although the LOAEL for atrazine seems high compared to other pesticides studied, the NOAEL is 3,500 μg/kg adipose tissue/day [95]. In an epidemiological study conducted in Spain between 1999 and 2009, high levels of DDT were found in the breast tissue of women who self-reported they were diagnosed with breast cancer. Among a total of 2,661 cases of breast cancer reported in this female study population, 2,173 (81%) were observed in areas of high DDT and atrazine contamination; leading to the conclusion that both pesticides are potentially correlated to the development of breast cancer [80]. Atrazine is associated with promoting mammary gland malignancies, especially when exposure to women occurs with the compound preferentially accumulating in adipose tissue of the breast, resulting in increased estrogen levels outside of the homeostatic range [81]. In a two-year study, rats were administered atrazine in their diet at varying concentrations in food at the ppm level. In females, after dosing of 500 or 1000 ppm, effects observed included bone marrow changes and related mammary adenocarcinomas and fibroadenomas [95].

Contrary to the aforementioned chemicals, phytoestrogens such as soy isoflavones, are directly ingested from plant-based foods; warranting the need for further study of examining various dosages and their potential effects on the human body. The amount of soy isoflavone consumption differs greatly not only amongst humans of different cultures, but also amongst those of the same ethnicity. For example, there is a 9-fold difference between daily intake of soy isoflavone intake consumed by Chinese-

Americans (4 g/day) compared to Chinese natives (36 g/day) [96]. These differences make it difficult to study the effects of soy isoflavone intake and what amount may pose risk of increased cancer incidence. Therefore, collectively, soy isoflavone intake has shown to impose adverse effects on the body at levels as low as 300 µg/kg bw/day [53] (**Table 5**). Studies conducted *in vitro* showed abnormal growth and proliferation after 100 nM injection of genistein on MCF-7 cells, as well as in mice with MCF-7 tumors in mammary glands [51]. Similarly, studies testing effects of daidzein found increased amounts of tumor invasion, promotion and proliferation in breast cells as well as weak binding to both estrogen receptors (Erα and ERß) [51, 96].

3.2 Mechanistic Pathways

3.2.1 Endogenous estrogen normal physiology

Estradiol (E2) undergoes major fluctuations throughout a woman's lifetime during the succession of pre-pubertal, menstrual, and menopausal stages [97]. This hormone is ideally kept within a range of 0.02-0.3 µg/L blood in adolescent girls and 0.03 – 8.0 µg/L blood in menstruating women [97]. In the genomic pathway, estrogen binds to the ER and induces conformational changes that accelerate DNA binding capability [83]. In the extra-nuclear pathway, rapid estrogen signaling occurs within seconds of addition of E2 and is mediated by the ERs that are localized to the cell surface membrane [83]. Estrogen response in the breast tissue is mainly considered to be mediated by estrogen-receptor alpha (ERα), stimulating cell proliferation, communication, apoptosis, etc. [98]. Adipogenesis and adipocyte differentiation is regulated in response to this estrogen-mediated mechanism, and therefore any kind of

dysregulation or disruption can lead to malformations [77]. EEDs have structures that are similar to steroid hormones such as estrogen, giving them the ability to activate or antagonize estrogen action at the ER in breast tissue [83] (**Table 6**).

Table 6. Mechanism of action of multiple estrogen mimicking endocrine disruptors at the estrogen receptor in breast tissue.

EED	Mechanism	Sources
BPA, daidzein, DDT, genistein	Estrogen Mimicry	[53, 83, 99]
Atrazine, daidzein, genistein	Estrogen Antagonist	[99] [53] [44] [80]
BPA, daidzein, DDT, genistein	Estrogen Agonist	[99] [53] [83] [80]

EED: Estrogen-mimicking endocrine disrupting chemical

3.2.2 Disruptive pathways of EEDs

Once BPA is ingested, its conjugated and unconjugated forms enter the metabolic pathways which then circulate in the bloodstream where the contaminant eventually interacts with the ER in breast tissue [79] working as either an agonist or antagonist. Typically, gonadotropin-releasing hormone (GnRH) is released by the hypothalamic-pituitary-ovarian axis and regulates luteinizing hormone and follicle-stimulating hormone, two hormones heavily involved in menstruation, and both of which play a major role in breast tissue development [83]. With BPA attached to the ER instead of estradiol, this could cause a disruption in the typical cycle of maturation and development, leading to mutations. Fluctuations of these hormones can also cause irregular cycles, as shown in an animal study conducted on adult mice where in utero exposure to BPA increased the duration of the estrous cycle [83].

Atrazine has the ability to increase the level of estrogen in the body by inducing aromatase activity [80]. Aromatase is an enzyme that converts testosterone to estradiol. Through this mechanism, estrogen builds up in the body, potentially causing adverse developmental effects, specifically in breast tissue [76]. Increased levels of estrogen which exceed the typical threshold for a menstruating woman, in the case of atrazine, has been linked to cancer of the mammary gland [81]. DDT is a strong estrogen mimic that has been linked to several abnormalities due to early exposure, such as premature menarche as much as 3-4 years earlier. Earlier menarche can lead to abnormal fluctuations of estrogen and other reproductive hormones necessary for breast development and therefore increases the risk of breast cancer later in life due to mutations in the tissues [76].

Genistein and daidzein both appear to act directly on the ER due to their structural similarities to estradiol [51]. This similarity allows them to act as a ligand, blocking the interaction between ER and estradiol, and activate transcription [53]. Genistein tumor effects occurred through its unique binding pattern with ERα where it binds with ER and triggers estrogen-dependent genes for regulating breast cancer, promoting metastasis [51]. ERα is influenced by genistein and daidzein as demonstrated in study where the presence of these compounds induces estrogen-receptor beta (ERβ) [53]. Daidzein is unique in that when it is digested in the large intestine, it is converted to its metabolite, equol [98]. Equol is a non-steroidal anti-androgen, therefore promoting more estrogen to accumulate inside the body [92]. Interestingly, only 25% of non-Asians and 50% of Asians possess the bacteria necessary to perform this conversion, indicating that some individuals may be more likely to benefit from soy consumption than others.

Additionally, equol production is largely varied amongst individuals and highly dependent on the current status of intestinal microflora [99].

Phthalates uphold lipophilic characteristics which strike a concern when studying breast cancer since adipose tissue predominantly makes up the breast anatomy [83]. While traveling throughout the body after ingestion, studies suggest that the phthalate molecule could accumulate in breast tissue and cause modification of mammary gland growth [54]. In addition to phthalates, parabens have been found to be more potent when administered through dermal exposure, studies have shown that long-term exposure, regardless of exposure route, increases the migratory and invasive properties of human breast cancer cells *in vitro* [83]. Due to this effect, parabens should be more closely investigated, especially when examining ingestion routes from dietary intake.

3.3 Conclusions

Exposure of women to multiple EEDs occurs on a daily basis throughout their lifespan, with dietary sources playing a prominent role. Although daily intake doses are low, the chronic nature of these exposures makes them last for a lifetime and can cause detectable steady-state levels of EEDs even in compounds that are fairly rapidly metabolized. These life-long, chronic exposures to substances of estrogenic potential pose notable threats to the human body and have been associated with an increased risk of breast cancer in cell culture [83], animal models [51] [96] [84], and in humans [83] [54]. It is evident that, based on the exposure route of dietary ingestion, compounds which pose the largest threat of breast cancer incidence are phytoestrogens, most notably the isoflavones genistein and daidzein as well as BPA. Isoflavones are consumed at

varied rates across ethnicities and dietary preferences, and also are fed at increased amounts to livestock consumed by humans [44, 53]. BPA exposures occur primarily as a result of migration of the chemical from food packaging and from the lining of cans into food; however, the compound also is found at a lesser amount in fresh fruits, vegetables, grains and other non-processed foods [79]. Although pesticide exposures are much smaller than those of BPA and phytoestrogens, there also is strong evidence that DDT and atrazine exert disrupting effects on physiological processes involving estrogen, thereby potentially leading to breast cancer [76, 86]. These findings suggest that a diet which emphasizes more fresh, unprocessed foods will reduce exposure to EEDs, and with it, potentially reduce the risk of breast cancer. Organically grown crops typically contain less unwanted agricultural chemicals and should be taken into consideration when purchasing foods at the grocery store [66]. In addition, this review identified a need for further efforts in examining the affinity of EEDs to the ER, in improving assessments of potency, and in studying daily exposures, windows of susceptibility and major exposure routes to inform the determination of exposure reference doses.

CHAPTER 4

UNDERSTAND POPULATION-LEVEL DYNAMICS OF DIETARY BEHAVIOR INTEGRATING A MULTI-OMIC APPROACH TO WASTEWATER-BASED EPIDEMIOLOGY

Introduction

In the previous two chapters, I explored the feasibility of expanding the use and application of wastewater-based epidemiology to other realms of public health beyond illicit drug monitoring was explored through the use of rigorous literature reviews with the goal of identifying potential new biomarkers indicative of nutritional status and chronic disease. Nutritional biomarkers were identified based on specificity of particular types of diet as well as connection to chronic disease. Detectability in urine and microbial biotransformation products from human gut microbiome interactions were also examined. Excreted biomarkers indicative of plant-based dietary behavior, such as the phytoestrogens genistein and daidzein, were also identified as potentially linked to breast cancer incidence; offering potentially preventative benefits. Additionally, utilizing this approach to be integrated into current public health surveillance programs had not been investigated more in-depth, allowing the opportunity to explore the potential benefits from a nutrition and human health perspective.

In Chapter 4, I conducted a two-year WBE study in a small catchment within a southwestern U.S. city, with the overarching goal to measure biomarkers indicative of diet and assess relevance to current measurements for nutritional assessment. Twenty-four-hour composite wastewater samples were collected for seven consecutive days each

month from within the sewer infrastructure and later analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS); compounds measured include genistein, daidzein, and enterolactone for the entire study period. Due to both genistein and daidzein being parent (ingested) compounds, a microbial metabolite of daidzein, equol, was later introduced in the study to test parent-metabolite interactions and determine suitability. Wastewater flow measurements and population estimates were used to calculate daily per capita phytoestrogen consumption and/or production for each individual compound and compared to existing estimates for average consumption of phytoestrogens in the U.S. Nonparametric statistical analyses were performed (Mann Whitney, Spearman's rank) to assess variability in consumption patterns between each year (year one, year two), seasons (fall, winter, spring, summer), months, and days (weekday vs. weekend). As proof of concept, interactions within the human gut microbiome were assessed by allocating a subset of samples for 16S rRNA sequencing as well as 16S rRNA quantitative polymerase chain reaction (qPCR) to determine microbial composition as it is relevant to the human gut, as well as test ability to detect and measure bacterial taxa that are reported to play a role in phytoestrogen metabolism.

4.1 Phytoestrogens and conventional methods of plant-based dietary assessment

Phytoestrogens are plant-derived, naturally occurring chemicals that have the ability to mimic estrogen, and/or interfere with the estrogen receptor. [100, 101] Due to this characteristic, phytoestrogens have received increasing attention in the clinical realm as an alternative to conventional hormone-related drug therapies. [94] More broadly, consumption of phytoestrogens suggests plant-based dietary behavior, which has been

shown to provide significant human health benefits, particularly for preventing nutrition-related chronic disease. [53, 94, 102] While survey-based methods have been employed to gain insight into dietary behavior, challenges experienced with these methods persist [16, 19-23, 103, 104]. For example, a 24-hour food recall survey offers the benefit of saving time, while still providing useful information on recent consumption. While it is considered the least biased of self-report methods, underestimating intake is possible due to the inability to capture the respondent's day-to-day variability in intake over long periods of time. [19, 20] Food diaries, in contrast, are designed to provide insight into dietary patterns over time, allowing the ability to make more robust associations between diet and disease. Administering this tool on various smartphone apps also makes it more accessible compared to other methods, however, ambiguity in describing preparation methods and under-/over-reporting can be common. It has been noted that in order to gain a comprehensive overview of dietary patterns and behavior, and depending on the study design and primary outcomes, multiple instruments may need to be employed, which may require a licensed professional to administer and analyze, such as a Registered Dietitian (RD), contributing to elevated study costs. [21, 105]

4.1.1 Wastewater-based epidemiology as a tool for measuring population-level diet

Wastewater-based epidemiology, a rapidly evolving scientific discipline where human excreted chemical and biological signatures are analyzed in community wastewater, offers a unique perspective into population-level behavior, exposure, and activity. Historically implemented at an international scale to monitor illicit drug use across cities, [5, 6, 27, 56] WBE has gained much more attention in recent years for its potential to expand into other realms of public health. [26, 37, 59]. The COVID-19 global

pandemic also highlighted the ability for WBE to transcend disciplines, with many groups now investigating viral and biological presence beyond SARS-CoV-2 at multiple scales of sample collection resolution [32-35, 43, 106-108]. While these studies have proposed dietary behavior as an opportunity for further exploration, few have fully investigated this conceivable venture, and none have examined this from a multidisciplinary analytical approach. As more evidence supports the need for alternative approaches to understanding dietary behavior due to the limitations mentioned above, and with many life-threatening diseases and conditions being linked to nutritional habits, it appeared prudent to explore this avenue of WBE to evaluate its utility in assessing population-level nutritional behavior to support public health interventions.

4.1.2 Dietary sources and cultural differences in consumption patterns

The most commonly consumed classes of phytoestrogens in the Western diet are currently understood to be isoflavones, lignans, and to a lesser extent, coumestans. A unique feature of phytoestrogen consumption as a whole are the observed cultural differences in consumption rates as it relates to breast cancer incidence. For instance, the mean daily phytoestrogen intake, predominantly through isoflavones, in Asian cultures can generally range between 25 – >50 mg d^{-1}, whereas in the United States, the per capita intake per day ranges between 1 – 3 mg d^{-1}, however, breast cancer incidence is markedly lower in the former population. [109, 110] Isoflavones, genistein and daidzein as the two primary compounds, are the major family of phytoestrogens that make up soy-based foods and food products, such as tofu, tempeh, fermented soybeans (natto), and soymilk, but are also found in an assortment of fruits and vegetables as well as processed foods at lower concentrations (**Appendix A; Table 8**) [111, 112]. Lignans are much more

ubiquitous in terms of food sources, including nuts, seeds, berries, beer, and a wide range of vegetables, with four major lignans commonly consumed: Lariciresinol, Pinoresinol, Matairesinol, and Secoisolariciresinol. The greatest concentration of lignan is known to be derived from flaxseed in the form of matairesinol and secoisolariciresinol, however, flaxseed is not commonly consumed in more Westernized cultures [113-115]. These consumed lignans are understood to eventually be transformed via microbial communities within the human gut and produce enterolignans, enterodiol and enterolactone; the latter being the final end-stage product of both human and microbial metabolism as enterodiol can further convert to enterolactone, but not vice versa [114, 115].

4.1.3 Human gut microbiome interactions

Daidzein has specifically been reported to convert to a more estrogenically active compound, equol, post-consumption through microbial interactions within the human gut. However, it is noted that approximately 25-30% of a Westernized human population (50-60% Asian population) is understood to contain the microbes necessary for this conversion, with so-called “non-producers” to yield *O*-desmethylangolensin (*O*-DMA), also through microbial interactions [116]. Equol has repeatedly been documented to be associated with reduced breast cancer incidence, and is also proposed as an alternative for hormone replacement therapy as well as symptom relief due to menopause. Some studies have also found positive associations with increased intake of isoflavone-rich foods and increased equol production as measured in urine [41, 116-118]. For enterolignans, specifically enterolactone, reports have noted increased production, measured through elevated urinary output, to be associated with cancer-protective properties as well as

support prevention of cardiovascular disease. Reports have also found an increase in use of oral antibiotics was associated with decreases in enterolactone in human blood and urine, indicating the measured values are predominantly driven by gut microbes [119]. Bacterial genera unique to the human gut microbiome and are currently known to play a role in either a direct or intermediate conversion to producing equol and enterolactone include: *Ruminococcus, Streptococcus, Eggerthella, Bacteroides, and Bifidobacterium*, amongst others [113, 120-122]. It has also been observed that an adoption of a vegetarian or vegan diet can promote growth of these beneficial bacteria in the gut [123]. These observed individualized human interactions and associated human health benefits warrants further exploration at population-scale to assess these links between diet and disease.

Thus, the goals of this study were to (*i*) investigate longitudinal trends of consumption of daidzein and genistein, and enterolactone production in a small sewer catchment to determine feasibility and appropriateness for a wastewater-based epidemiology study, (*ii*) assess microbial composition of biologically-relevant taxa known to play a role in phytoestrogen consumption and human health, (*iii*) explore the potential of novel biomarkers, such as equol, to enhance insights into isoflavone consumption and subsequent gut microbial production, and (*iv*) discuss opportunities for future investigation.

4.2 Materials and Methods

4.2.1 Chemicals and reagents

Native analytical standards of genistein, daidzein, and enterolactone were purchased from Sigma-Aldrich (St. Louis, MO, USA). Isotopically labeled analytical standard genistein-*d4* was purchased from Cayman Chemical (Ann Arbor, MI). Native analytical standard of equol was purchased from Santa Cruz Biotechnology (Dallas, TX, USA). Liquid chromatography-grade (LC-grade) water, methanol, and acetone were obtained from Fisher Scientific (Waltham, MA, USA), and LC-grade formic acid was purchased from Fluka Chemical Corp (Milwaukee, WI, USA). Stock standard solutions were prepared in-house using LC-grade methanol and stored at -20℃. Temporary working solutions were prepared by serial dilution of stock solutions with LC-grade methanol and stored at -20℃. All glassware used for this study was washed with laboratory-approved detergent, rinsed with LC-grade water, capped with aluminum foil, and heated at 550℃ for 4 hours with a 12-hour cooling period prior to use.

4.2.2 Study location

This study took place in a small sub-catchment of a large urban city located immediately south-east to a major public university. This catchment, at the time this study took place (August 2017 – July 2019), was comprised of a mix of single-family homes, condominiums, apartment complexes, off-campus student housing, a public park, restaurants, and hotels. Demographically, the area is comprised of White Non-Hispanic (~40%), Asian (~25%), Hispanic or Latino (~21%), African American (~7%), and American Indian/Alaska Native (~4%) individuals. The median household income was approximately $37,000, with an average household size of 3 and median age of 24 years

old. The wastewater collection system drains south-west, with sewage retention times ranging between 2 (minimum) and 50 (maximum) minutes, with an average time of approximately 20 minutes. Documented historical estimates of wastewater temperatures unique to this catchment at time of collection ranged between 20℃ – 30℃ depending on time of year.

4.2.3 Sample collection and transport

Time-weighted, twenty-four (24) hour composite wastewater samples were collected over seven (7) consecutive days per month between August of 2017 and July of 2019 (n = 156). An Avalanche® automated refrigerated sampler (Teledyne, ISCO, Lincoln, NE, USA) deployed above ground and maintained inside a permanent cabinet station was set to collect 60-100 milliliters (mL) of raw wastewater from within the sewer collection system every fifteen minutes; captured in previously acid-washed 10 L glass vessels. Each 24-hour composited sample was adequately mixed and transferred to two-liter high-density polyethylene (HDPE) bottles; immediately placed on ice in designated coolers and transported to the nearby laboratory. Typical 24-hour sample collection timeframes occurred from 7:00AM-6:59AM; time between sample collection to same-day processing did not exceed 1 hour. Wastewater flow measurements were monitored by ISCO LaserFlow flow meters (Teledyne, ISCO, Lincoln, NE, USA), located within the permanent sample collection station which monitored and recorded flow at 2-minute intervals. Flow measurements were obtained through FlowLink online software (Teledyne, ISCO, Lincoln, NE, USA).

4.2.4 Sample processing and analysis

Methods employed for chemical processing and analysis have been previously published. [39, 124] Briefly, duplicated 200 milliliter (mL) aliquots of each raw influent wastewater sample were spiked with isotopically-labeled internal standard (1 μg L^{-1} final concentration in sample) and subsequently arranged on a Dionex™ Autotrace™ 280 Solid-Phase Extraction Instrument (SPE) (Thermo Scientific, Waltham, MA) using Oasis Hydrophilic-Lipophilic Balance (HLB) cartridges (150mg, 6cc, 30 um particle size) (Waters, Milford, MA). Method blanks (deionized water) were extracted and analyzed alongside each set of samples to determine potential contamination. Prior to sample loading, cartridges underwent conditioning with methanol, followed by a water rinse. Post-sample loading, cartridges were dried under a gentle stream of nitrogen for 10 minutes. Next, gravity drip-wise elutions were performed using a vacuum manifold with an in-line HEPA filter using a 1:1 (*v:v*) methanol and acetone solution with 0.5% formic acid until a final volume of 4mL was achieved. Organic extracts were then stored at -20℃ until further analysis.

For LC-MS/MS sample extract preparation, 200 microliters (μL) of each organic extract was aliquoted into glass amber vials with 350 μL inserts and dried down under a gentle stream of nitrogen. Extracts were reconstituted first with 100 μL of LC-MS grade methanol, followed by 100 μL of LC-MS grade water, and then lightly vortexed. Finalized extracts were analyzed for targeted analytes using a Shimadzu Prominence 2100 high performance liquid chromatographer (HPLC) (Marlborough, MA) paired to an AB Sciex API 4000 triple quadrupole mass spectrometer (Applied Biosystems, Framingham, MA) with electrospray ionization (ESI) operating in negative mode.

Analyte identification was achieved using compound-specific retention times and ion transition from multiple reaction monitoring (MRM) (Appendix A: Table 9). Chromatographic separation was attained using a Symmetry C_8 column (4.6 x 150 mm, 3.5 um particle size), followed by a Symmetry VanGuard Cartridge (3.9 x 5 mm, 3.5 um particle size) (Waters, Milford, MA). Further specific method details are listed in Appendix A: **Table 10**.

4.2.5 Quality assurance/quality control

Negative extraction blanks (deionized water) were incorporated with each SPE run alongside samples to identify potential contamination. Reported concentrations were determined based on a 14-point standard curve (daidzein, genistein, enterolactone), ranging from 0.05 μg L^{-1} to 2000 μg L^{-1}, and a 12-point standard curve (equol) ranging from 0.5 μg L^{-1} to 2000 μg L^{-1} with a minimum coefficient of determination of $R^2 = 0.99$. Precision was expressed as Relative Percentage Difference (RPD); target precision between replicates was <30% (Equation 1). Instrument blanks (50:50 methanol and water) were included every six to eight samples to assess analyte carryover, of which no carryover was observed in this study. Detailed method validation information including method detection limits (MDL) and recoveries are provided in Appendix A: Table 11. Chromatograms of select analytes can be found in SI (Appendix A: **Figure 17**). Relative Percentage Difference (RPD) calculated using the following equation:

$$RPD\ \% = ABS\left(\frac{C_{S1} - C_{S2}}{C_{S1} + C_{S2}}/2\right) \times 100 \qquad \text{Eq. 1}$$

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

4.2.6 Data analysis

LC-MS/MS data were acquired with Analyst 1.5 software (Applied Biosystems, Foster City, CA), where concentrations were calculated using isotope-dilution and subsequently reported if the signal-to-noise ratio was greater than 10 and concentrations were above the MDL. Calculated concentrations (μg L^{-1}) were converted to mass loads (g d^{-1}) using flow data provided by the municipality and accessed with FlowLink software (**Table 7**).

Table 7. Population estimates and flow measurements used for specific time periods throughout this study.

Time Period	Population	Flow (Liters/day)
August – May	9,848	2,165,424
June – July	6,976	1,636,932

Population-normalized mass loads (μg d^{-1} capita) were produced using population estimates as described below and listed in **Table 7**. Per capita daily genistein consumption (GC) was calculated using the following equation:

$$GC = \frac{C_G * Q_{Tot}}{CF_G * Pop} \qquad \text{Eq. 2}$$

Where C_G is the measured genistein concentration, Q_{Tot} is the total daily volumetric flow rate, CF_G is the correction factor (5; 20% excretion) [125, 126] (Appendix A: **Table 12**). *Pop* is estimated population. Per capita daily daidzein consumption (DC) was calculated using the following equation:

$$DC = \frac{C_D * Q_{Tot}}{CF_D * Pop} \qquad \text{Eq. 3}$$

Where C_D is the measured daidzein concentration, Q_{Tot} is the total daily volumetric flow rate, CF_D is the correction factor (2.2; 45% excretion) [126, 127] (Appendix A: **Table**

12). *Pop* is estimated population. Per capita daily lignan consumption (LC) was calculated using the following equation:

$$LC = \frac{C_{ENT} * \left(\frac{MW_{LIG}}{MW_{ENT}}\right) Q_{Tot}}{EF_{ENT} * Pop} \qquad \text{Eq. 4}$$

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 d^{-1}), previously reported in WBE studies (Appendix A: **Table 12**) [124], and *Pop* is estimated population. Per capita daily equol production (EC) was calculated using the following equation:

$$EC = \frac{C_{EQ} * \left(\frac{MW_{DAD}}{MW_{EQ}}\right) Q_{Tot}}{EF_{EQ} * Pop} \qquad \text{Eq. 5}$$

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 d^{-1}) [121] (Appendix A: **Table 12**), and *Pop* is estimated population.

Per person consumption was calculated using previously reported methods with slight modification. [128] Resident population estimates were determined using 2010 census block group data and employment data using Maricopa Association of Government (MAG) where resident and non-resident employment was examined. Student population estimates were obtained from publicly-available campus resident data and estimates using changes in wastewater flow volume (**Table 7**). [128]

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 (FDR): 0.05) using the following equation:

$$BH = \frac{i}{m} Q \qquad \text{Eq. 6}$$

Where i is the rank assigned to the p-value in the array, m is the number of comparisons, and Q is the FDR. Spearman rank-order correlations were performed to assess correlations between parent-metabolite (daidzein, equol) interactions as well as assess in-sewer degradation.

Effect of temperature on in-sewer analyte degradation was assessed through statistical analysis. Reported ambient temperatures (average, minimum, maximum) for each sample collection day throughout this study were recorded. A Spearman's rank order non-parametric test was first used to compare the recorded daily ambient temperature with estimated wastewater temperature previously reported (**Appendix A: Table 13**; **Figure 16**) [129] based on historical estimates within the study area. Next, the relationship between recorded ambient temperatures and measured analyte signal in wastewater (μg/L) for each analyte (genistein, daidzein, enterolactone, and equol) was also assessed using Spearman's rank order non-parametric tests ($\rho < 0.50$ = weak; $\rho > 0.50 < 0.70$ moderately strong; $\rho > 0.70$ strong).

4.2.7 Microbiome analysis

As a proof of concept for understanding human gut microbial interactions at population-level, a subset of samples (n = 12 months) was allocated for microbiome analyses. Approximately 84 previously frozen raw wastewater samples from the entire

year of 2018 (Jan through Dec; 7 samples/month) were thawed, gently inverted and well mixed, and aliquoted into 100 mL 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, Burlington, MA) using a vacuum pump apparatus with in-line filter, discarding the filtrate. Each filter was then aseptically transferred into an individual bead tube and underwent bead-beating for approximately 15 minutes using a Vortex Genie 2 with an adapter to secure the tubes (Scientific Industries, Bohemia, NY). DNA extractions then were performed using a QIAGEN DNeasy Power Soil Pro Kit (Hilden, Germany), following manufacturer's instructions. The extracted DNA (50 uL) was immediately stored at -80°C until further analysis. A whole process negative extraction control (deionized water) was incorporated to account for potential contamination throughout the extraction process.

Bacterial community composition analysis was performed with next generation sequencing in MiSeq Illumina platform. Amplicon sequencing of the V4 region of the 16S rRNA gene was performed with the barcoded primer set 515f/806r [130], following the protocol by the Earth Microbiome Project (EMP) (www.earthmicrobiome.org) for library preparation [131]. PCR amplifications for each sample were performed in duplicate, then pooled and quantified using the Accublue® High sensitivity dsDNA Quantitation Kit (Biotium, Fremont, CA). A no template control (NTC) sample was included during the library preparation as a control for extraneous nucleic acid contamination. 200 ng of DNA per sample are pooled and then cleaned using a QIA quick PCR purification kit (QIAGEN, Hilden, Germany)). The pool was quantified by Illumina library Quantification Kit ABI Prism® (Kapa Biosystems, Wilmington, MA).

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 in the MiSeq Illumina and run using the version 2 module (2 x 250 paired-end) following the directions of the manufacturer.

Data analysis was achieved using QIIME 2 (version 2021.2) [132] for sequence quality control and feature table construction. The DADA2 plugin [133] was used to filter and merge the forward and reverse reads. Sequences were then mapped to the Silva 138 SSURef NR99 Database 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 using the R phyloseq package (version 1.36.0) [134], and contaminating ASVs were identified and removed with the R decontam package (version 1.12.0) [135]. ASVs were classified as contaminants if identified by either the frequency or prevalence methods (method = "either"). Multiple probability thresholds were tested: 0 (no contaminants), 0.1 (default threshold in decontam), through to 0.5, the most aggressive threshold considered. A final threshold of 0.5 was chosen for data interpretation due to the low biomass nature of the samples and reduce likelihood of interpreting a contaminant as a viable sequence. 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). [136]

Quantitative polymerase chain reaction (qPCR) was then performed using a QuantStudio 3 (Applied Biosystems, ThermoFisher Scientific, Waltham, MA) instrument to quantify the total 16S rRNA gene as a proxy of bacterial concentration (16S rRNA gene copies L^{-1} wastewater) in each sample. A previously published and validated

TaqMan-based assay was employed targeting the 16S rRNA gene using a universal primer and probe set [137] (**Appendix A**: **Table 14**). All samples were run in triplicate along with a standard curve (16S rRNA plasmids) ranging from 10^7 to 10^1 copies uL^{-1} along with molecular no-template controls (NTC) using RNAse/DNAse-free Ultrapure™ PCR-grade water (Invitrogen, Waltham, MA). Each reaction contained 300 nM probe and 300 nM of each primer, 1X concentrated buffer (TakaraBio Inc, Kusatsu, Shiga, Japan), 2 uL DNA template, and nuclease-free water up to 20 uL final reaction volume. Thermal cycling conditions were as follows: Hot start at 95℃ for 2 minutes, followed by 45 cycles of 95℃ for 10 seconds, 56℃ for 20 seconds, and 68℃ for 20 seconds.

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

4.3 Results and discussion

The goal of this study was to investigate the feasibility of conduct WBE in a sub-catchment of a southwestern U.S. city from August 2017 through July 2019 to understand population-level dynamics of plant-based dietary behavior. A multi-omics approach was used in order to gain a comprehensive understanding of dietary consumption patterns in a community of interest (**Figure 2**).

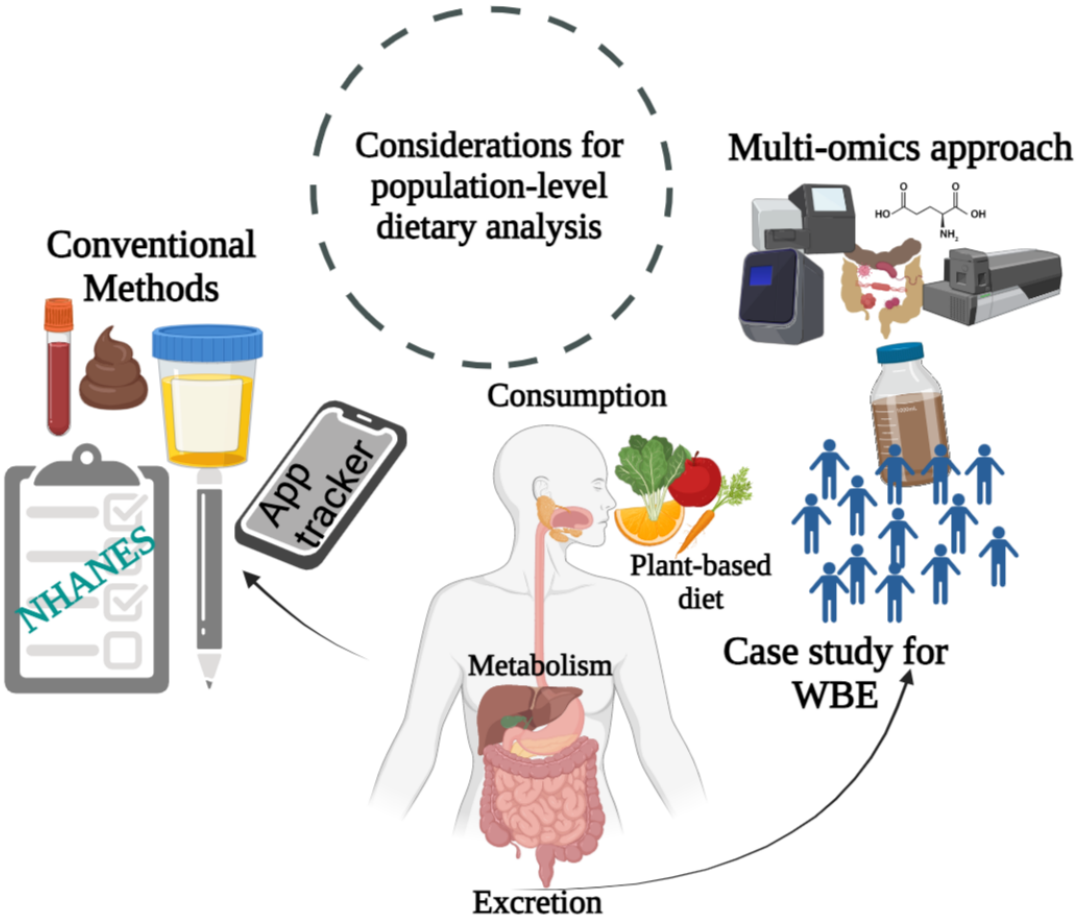

Figure 2. Visual representation of the various methods and protocols considered and/or employed in this study for understanding population-level dynamics of dietary behavior. A two-year case study conducted investigating plant-based dietary behavior utilizing wastewater-based epidemiology (WBE) to measure phytoestrogens through a multi-omics analytical approach that serve to inform considerations for future work. Conventional methods (not used here) include: Individualized human biomonitoring, personal app trackers on smart phones, and survey methods such as 24-hour food recall or national surveys such as National Health and Nutritional Examination Survey (NHANES). Created using BioRender.com.

Seven sample collection days per month were determined by the city based on compliance monitoring; with a total of 168 potential sample collection days. Of those, 12 samples were lost due to automated sampler malfunction or other related event (i.e., clogging of tubing) leaving a final total of 156 samples collected, processed, and analyzed; with 100% detection for all phytoestrogens under investigation. Measured

concentrations in raw wastewater, calculated mass loads, and average per capita excretion are shown for isoflavones genistein (**Figure 3A**) and daidzein (**Figure 3B**), along with enterolactone (**Figure 4A**) and equol (**Figure 4B**). Calculated average per capita mass loads of genistein, daidzein, enterolactone, and equol (avg ± SD μg d^{-1} per capita) were 255 ±106 μg d^{-1} per capita, 871 ±553 μg d^{-1} per capita, 1437± 751 μg d^{-1} per capita, and 231± 142 μg d^{-1} per capita, respectively. Measured per capita mass loads reported herein are higher in relation to previous WBE studies of measuring genistein, daidzein, and enterolactone in raw wastewater: 94± 30, 310 ±79, 349± 57 μg d^{-1} per capita (City 1) and 108 ±45, 490± 216 645± 265 μg d^{-1} per capita (City 2) [39], however, this could be due to sample collection from within the sewer infrastructure as employed here, compared to a wastewater treatment plant (WWTP) which could increase risk for analyte degradation due to long in-sewer retention times.

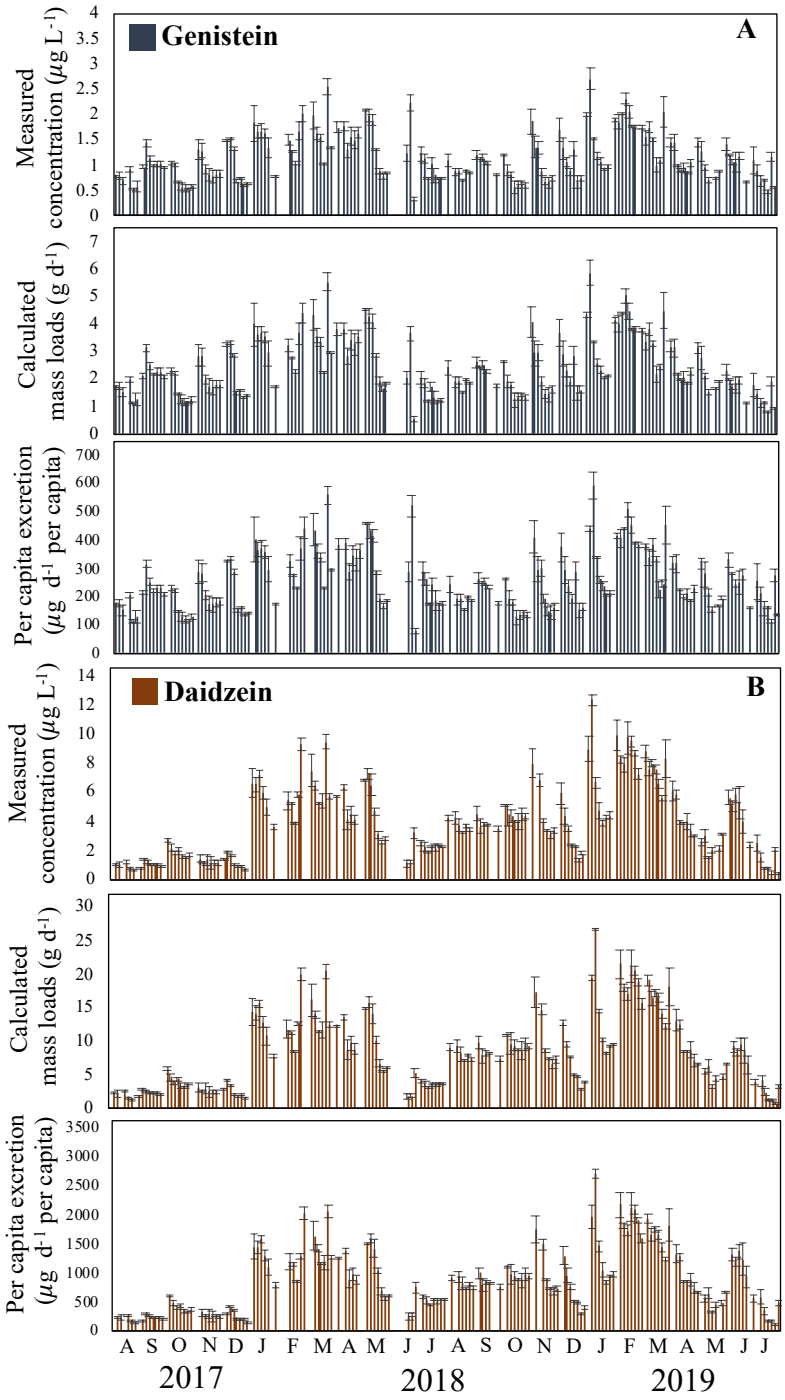

Figure 3. Measured average concentrations in raw wastewater (μg L^{-1}), calculated mass loads incorporating flow estimates (g d^{-1}), and per capita excretion (μg d^{-1} per capita) for isoflavones **(A)** genistein and **(B)** daidzein for the entire study period beginning August 2017 through July 2019. Error bars represent standard deviation between duplicates.

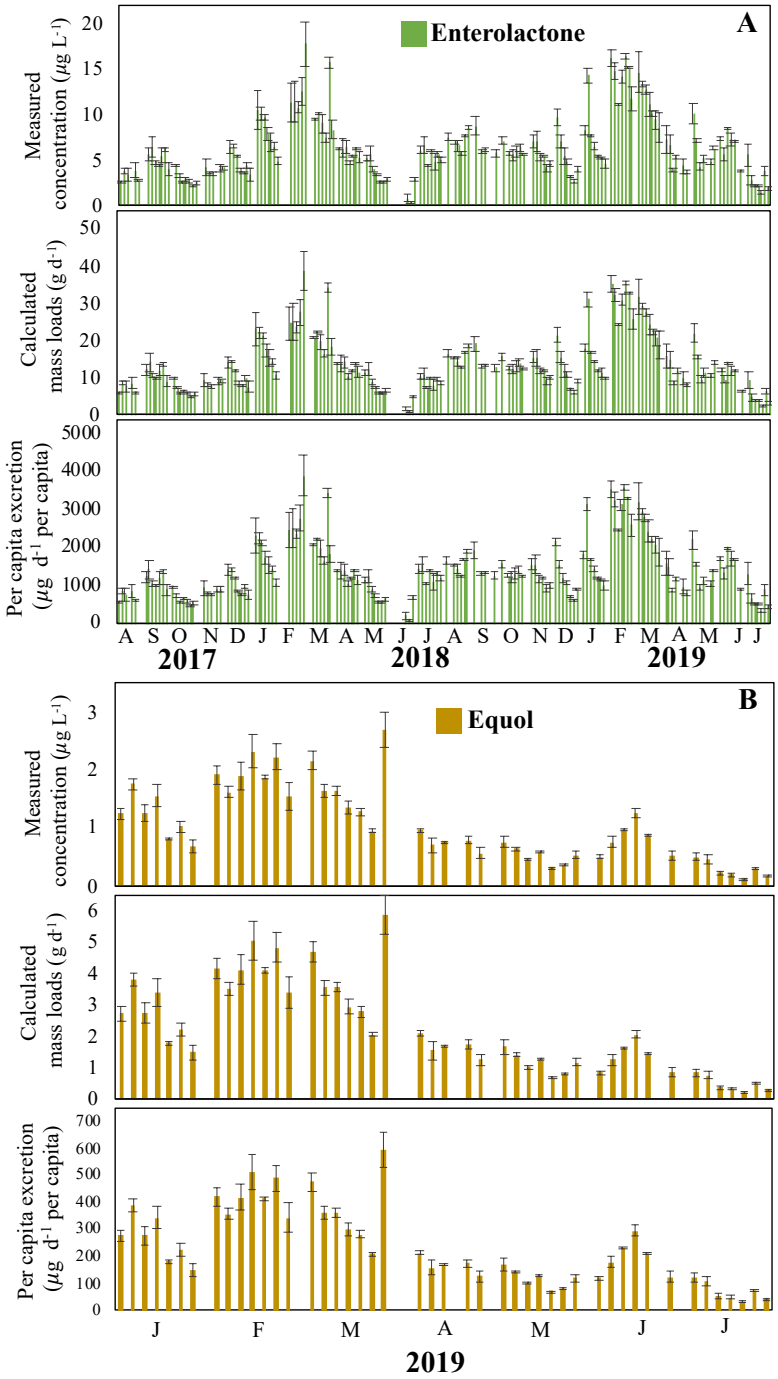

Figure 4. Measured concentrations in raw wastewater (μg L^{-1}), calculated mass loads incorporating flow estimates (g d^{-1}), and per capita excretion (μg d^{-1} per capita) for **(A)** enterolactone (August 2017 through July 2019) and **(B)** equol (January 2019 through July 2019). Error bars represent standard deviation between duplicates.

In the first year of monitoring, total phytoestrogen consumption patterns displayed a distinct increase at the start of each new year with a subsequent decline after the month of March. This pattern was repeated in the following year (**Figure 5A; Appendix A: Figure 16**). This increase at the beginning of the year could be due to lifestyle changes with increased purchasing and consumption of healthier foods after making New Years' resolutions [138]. Comparisons between years (year one vs. year two), days of the week (week vs. weekend), and seasons (Fall, Summer, Spring, Summer) were assessed, with statistically significant differences between year one and year two ($p \leq 0.01$), and increased consumption overall in the second year (**Figure 5B**). While no statistically significant difference between weekday and weekend consumption was observed, seasonal changes specifically between Fall and Winter and Spring and Summer held statistical significance ($p \leq 0.01$) (**Figure 5C-D**). Seasons were determined according to the National Geographic Society [139]: Fall (September, October, November); Winter (December, January, February); Spring (March, April, May); Summer (June, July, August). These changes are in agreement with the overall consumption patterns across the year, with large increases occurring at the beginning of the year (winter) and declining after March heading into the summer.

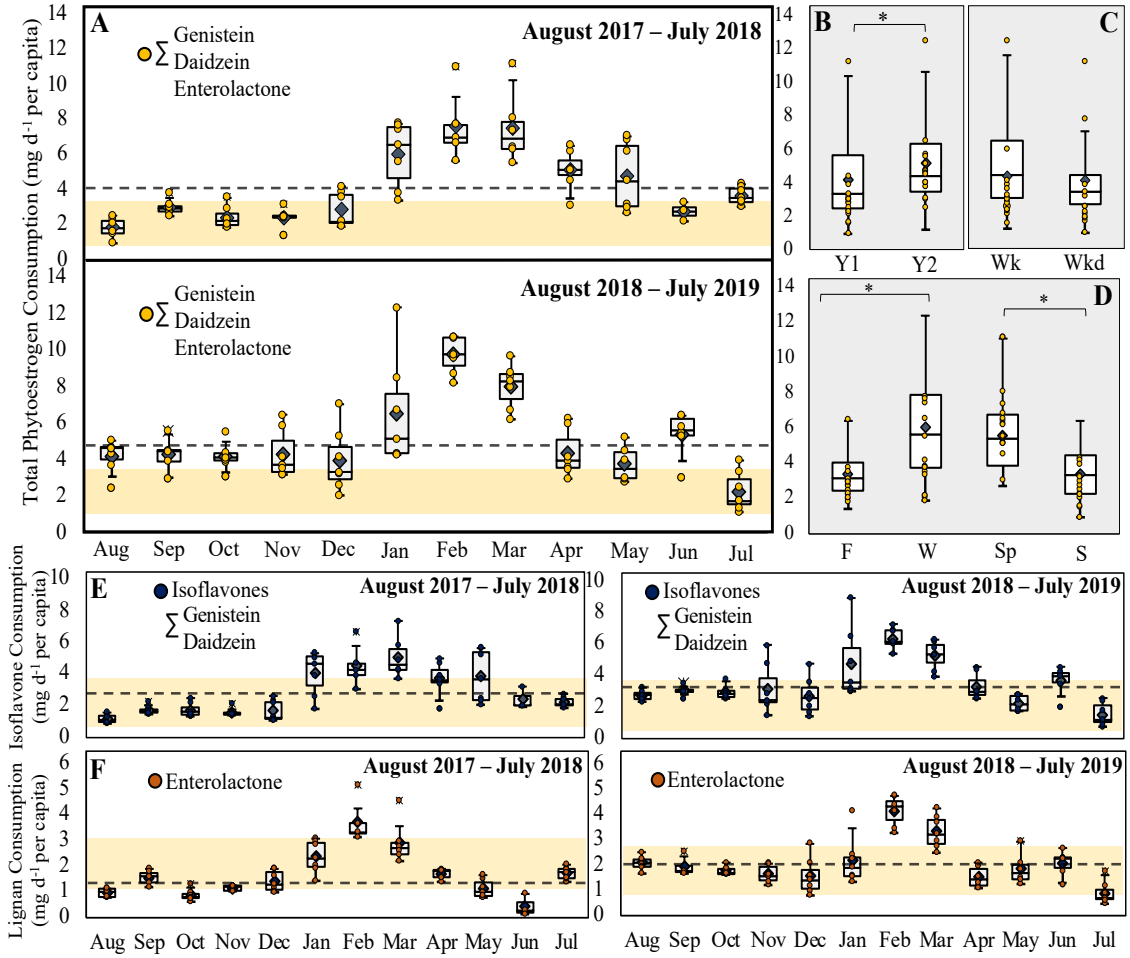

Figure 5. Calculated per capita consumption of phytoestrogens based on measurements from August 2017 through July 2019 all in **mg d^{-1} per capita.** Thick dashed lines represent measured averages for each analyte or group of analytes in this study. Yellow shading represents literature reported estimated range of phytoestrogen consumption in a western society (1-3 mg d^{-1} per person). Box plots show the 25th, 50th (median), and 75th quartiles with minimum and maximum error bars. Black diamonds represent the mean for each month and yellow circles are each individual points. **(A)** Total phytoestrogen consumption (sum of genistein, daidzein, and enterolactone) measured in raw wastewater **(B)** Year by year comparison for total phytoestrogen consumption showing statistically significant changes **(C)** Weekday (Wk) versus weekend (Wkd) comparison for total phytoestrogen consumption and **(D)** Seasonal comparison showing total phytoestrogen consumption in Fall (F), Winter (W), Spring (Sp), Summer (S) with statistically significant changes between F and W and Sp and S. **(E)** Isoflavone consumption as the sum of genistein and daidzein for both years. **(F)** Lignan consumption indicated by the production of the microbial metabolite enterolactone for both years. Statistical significance informed by Mann Whitney *U* nonparametric test of variability with Benjamini Hochberg (BH) correction for false discovery (0.05) (*p≤0.01).

As previously mentioned, it is estimated that the average consumption rate of phytoestrogens in a western society ranges between 1-3 mg d^{-1} per person, however, these estimates can vary depending on life stage, dietary lifestyle preference, and others such as demographics or food access. [101, 140] Average total phytoestrogen consumption was slightly higher than this estimated range; year one 4.1$\pm$ 2.2 mg d^{-1} per capita and 5.0 $\pm$ 2.3 mg d^{-1} per capita for year two. However, these estimates reflect dietary behavior of a western society, which has decreased representation from minority groups than in this study. For example, according to the U.S. Census Bureau, demographic populations 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. [141] This difference in demographic distribution could result in the varied consumption rates. Measured isoflavone and lignan consumption informed by wastewater, however, was consistent within the 1-3 mg d^{-1} estimated range, although the isoflavones were closer to the higher end (**Figure 5E**), which could be due to the greater percentage of Asian residents within this catchment. Lignans, as described, are much more ubiquitous than isoflavones, and found in a wide variety of fruits and vegetables as well as beer. [115] Lignan consumption was represented by measuring enterolactone, an end-product microbial metabolite of consumed lignans (**Figure 5F**). [113, 142] Per capita consumption and/or production (mg d^{-1} per capita) for the entire study period for genistein, daidzein, and enterolactone can be found in **Figure 6A**. Statistical assessment to determine susceptibility to temperature degradation tested by Spearman's rank order nonparametric

test are displayed in **Figure 6B**. Full details of associated *p*-values and Benjamini-Hochberg (BH) corrections for each analyte are shown in Appendix A: **Table 15**.

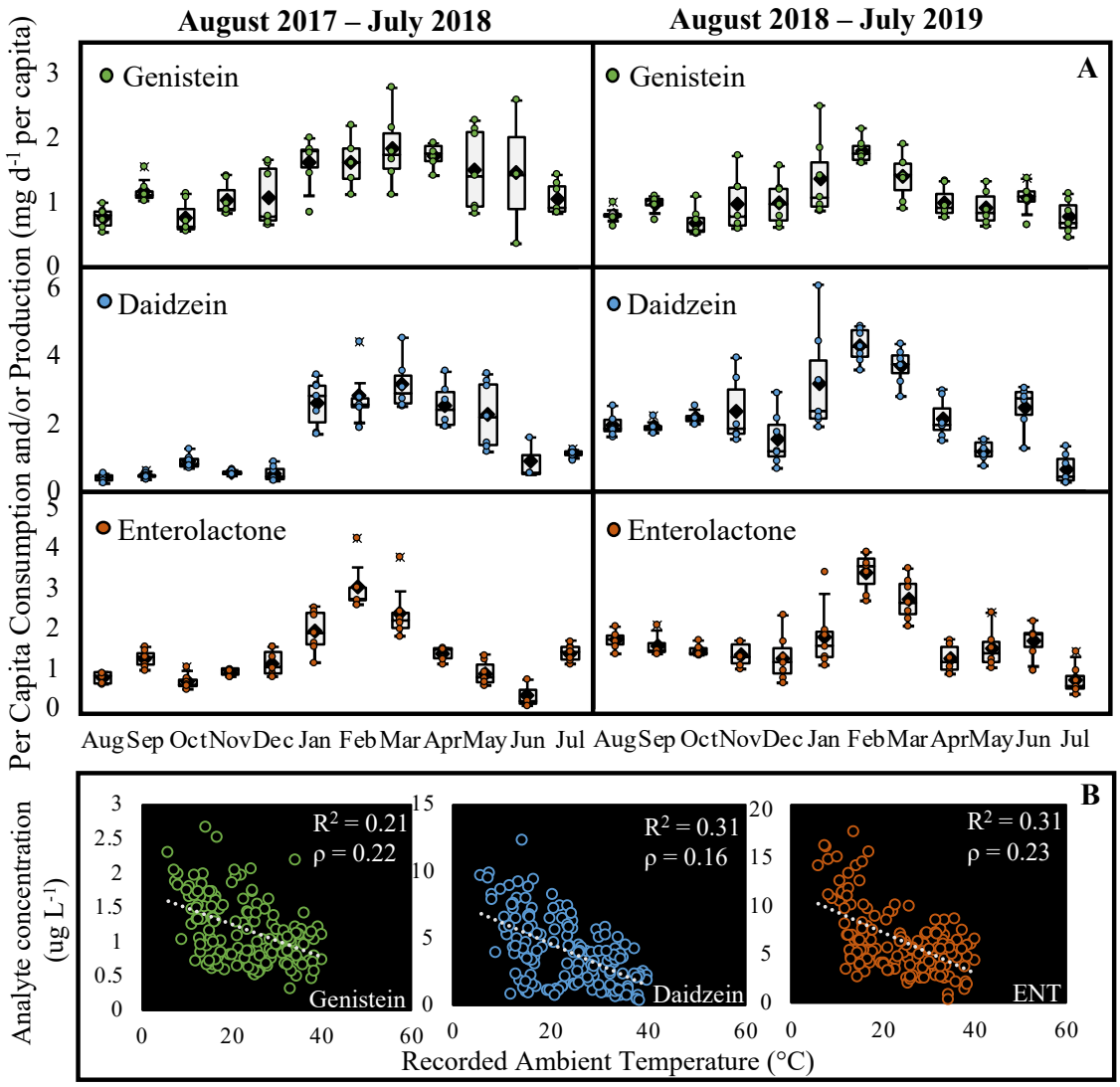

Figure 6. (A) Per capita consumption (genistein, daidzein) and production (enterolactone) shown in mg d^{-1} per capita from August 2017 through July 2019 and **(B)** Spearman's rank order nonparametric results comparing measured concentration for each analyte (μg L^{-1}) with recorded daily ambient temperature for each sample collection day throughout this study. Tests resulted in weak associations (ρ<0.50) for genistein (ρ = 0.22), daidzein (ρ = 0.16), and enterolactone (ρ = 0.23).

Phytoestrogens are known to have varying degrees interactions within the human gut microbiome interactions at the individual level. To investigate potential relationships at population-scale, a subset of samples (n = 84; January through December 2018) was allocated to test if these interactions identified in the literature are reflected in community municipal wastewater. Compositional analysis down to the genus-level of each month revealed that specific bacterial taxa that have been isolated from within the human gut microbiome, such as *Bifidobacterium.*, *Blautia*, and *Romboutsia*, were detectable in a community sewage sample (**Figure 7A**). In more recent studies, these genera have also been reported to interact with phytoestrogens and produce metabolites such as enterolactone and equol [42, 118]. Few studies investigating at the individual-level have also observed anti-cancer properties of these microbial products, as well explore whether increases in certain genera would result in greater enterolactone production, begging the question of the role of diet in promoting growth of these beneficial microbes [143]. Using the measured relative abundance and total 16S rRNA genes (a proxy for bacteria concentration as 16S rRNA gene copies L^{-1} wastewater) (**Figure 7B**), average semi-quantitative abundances of the number of 16S rRNA genes belonging to select enterolignan- and equol-producing genera were calculated, including: *Ruminococcus* (1.56×10^{10}), *Clostridium* (5.42×10^{9}), *Bacteroides* (2.08×10^{9}), *Blautia* (1.20×10^{10}), *Bifidobacterium* (4.57×10^{10}), *Romboutsia* (1.65×10^{10}), *Intestinibactre* (1.12×10^{9}), *Subdoligranulum* (1.45×10^{10}), *Dorea* (9.78×10^{9}), *Slackia* (1.98×10^{7}), *Prevotella* (4.29×10^{8}), *Roseburia* (2.14×10^{7}), *Senegalimassilia* (8.66×10^{7}), and *Eubacterium* (1.36×10^{9}) (**Figure 7A**). 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%). Measured per capita consumption rates (mg d^{-1} per capita) are shown in **Figure 7C** to visualize trends between chemical and microbial data. Notably, the month with the least estimated total phytoestrogen consumption and/or production (June) determined in wastewater for the entire study period corresponds to the lowest 16S rRNA bacterial gene copies L^{-1} wastewater measured in June of 2018 determined by qPCR (**Figure 7C**). In general, it could be assumed that a lower amount of 16S rRNA bacterial gene copies L^{-1} could indicate a decreased abundance of bacteria in that given sample, leading to decreased production of measured enterolignans produced in the gut, however, similar trends were not observed when comparing the chemical and biological data for any other month under study. Thus, this phenomenon should be further explored as it could be utilized in predictive models to inform on anticipated consumption trends and related disease incidence patterns. These results suggest the feasibility of observing human gut microbiome at community-scale, and further, to interpret results in the context of population diet and health.

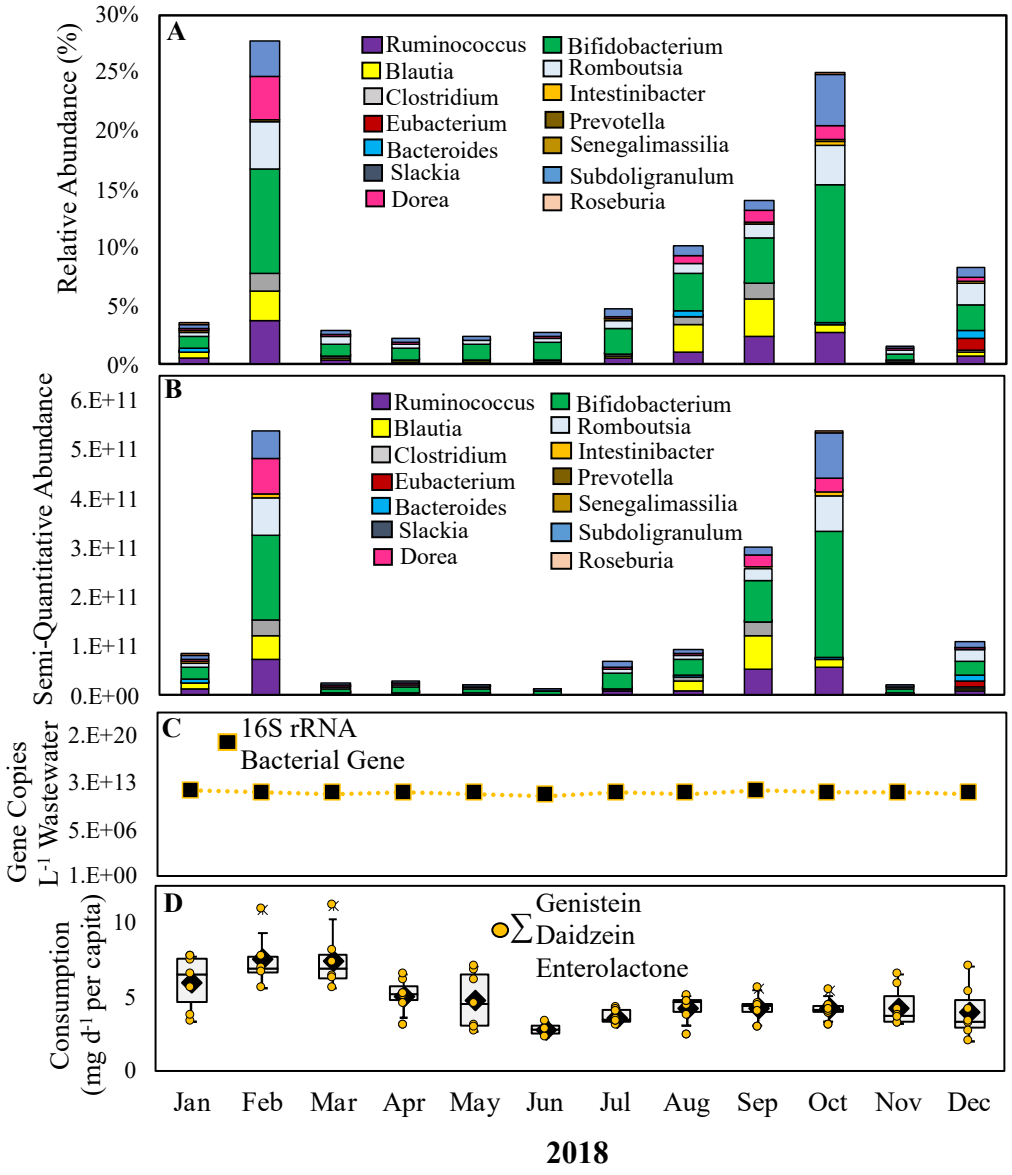

Figure 7. Microbial composition from a subset of wastewater samples investigated in this study (January through December 2018) **(A)** Relative abundance (%) of select bacterial taxa at genus-level involved in phytoestrogen metabolism within the human gut, (**B**) Semi-quantitative abundance of the same genera **(C)** Measurement of total bacterial 16S rRNA genes in each sample from quantitative polymerase chain reaction (qPCR) ((log) gene copies 16S rRNA L^{-1} wastewater) **(D)** Total phytoestrogen consumption (mg d^{-1} per capita) as the sum of genistein, daidzein, and enterolactone measured in raw wastewater. Box plots show the 25th, 50th (median), and 75th quartiles with minimum and maximum error bars. Black diamonds represent the mean for each month and yellow circles are each individual points.

It has been previously reported that measuring genistein and daidzein to understand human consumption could be a challenge as these are both parent (ingested) compounds that could also be measured as a result from industrial or restaurant dumping and/or runoff. [24] While this study was conducted from within the sewer network at a more granular level than a WWTP, it was important to test the efficacy of using these compounds as biomarkers for human consumption by incorporating a metabolite that could potentially be used for future study. In an effort to achieve these goals, equol, a microbial metabolite of daidzein, was included starting in January of 2019 and measured through July 2019 (**Figure 8**). It is understood that equol is produced by approximately 30% of populations who live in a western society, with some studies showing that an increase in chronic consumption of foods that contain daidzein can promote microbial growth amenable for equol production [120, 121]. Due to this distinct attribute, it calls to question how this should be interpreted in a WBE study. It is possible that using an average per capita estimate to assess equol production in a community would lead to underestimation, as shown in **Figure 8**. Thus, these measurements were corrected to account for one-third of the population to have the ability to produce equol, increasing the amount produced, and could be a more appropriate way to display equol production at population-level. Trends in consumption of daidzein and production of equol were then examined, again displaying equol as both a per capita average and 30% population corrected (**Figure 9A**). The trends observed here are in-line with the other phytoestrogens investigated in this study; elevated levels at the beginning of the year with a gradual decrease thereafter. Spearman's rank order nonparametric test indicates a strong

correlation between daidzein consumption and equol production (**Figure 9B**) ($\rho = 0.84$) with a daidzein to equol ratio of 0.2, corresponding closely to the 30% (i.e., 0.3) production rate of a Western society, and suggesting not only that the use of equol serves as an ideal candidate for future WBE investigation, but that the use of daidzein is valid as a dietary biomarker, depending on the location and resolution of sample collection. Overall, due to the selectivity of equol production amongst a population, the connection to dietary behavior to promote beneficial microorganisms in the human gut, and the associated degree of estrogenic potency of equol that offers potential human health benefits, it appears logical to pursue equol for future WBE investigation to determine population-level trends in production; later identifying links to associated chronic diseases such as breast cancer. Finally, due to these aforementioned qualities, use of equol as an alternative for hormone-replace therapy (HRT) as well as symptom relief for other hormone-related occurrences such as menopause has gained clinical attention [144]. This further strengthens the case to understand these trends using an objective, population-level assessment.

.

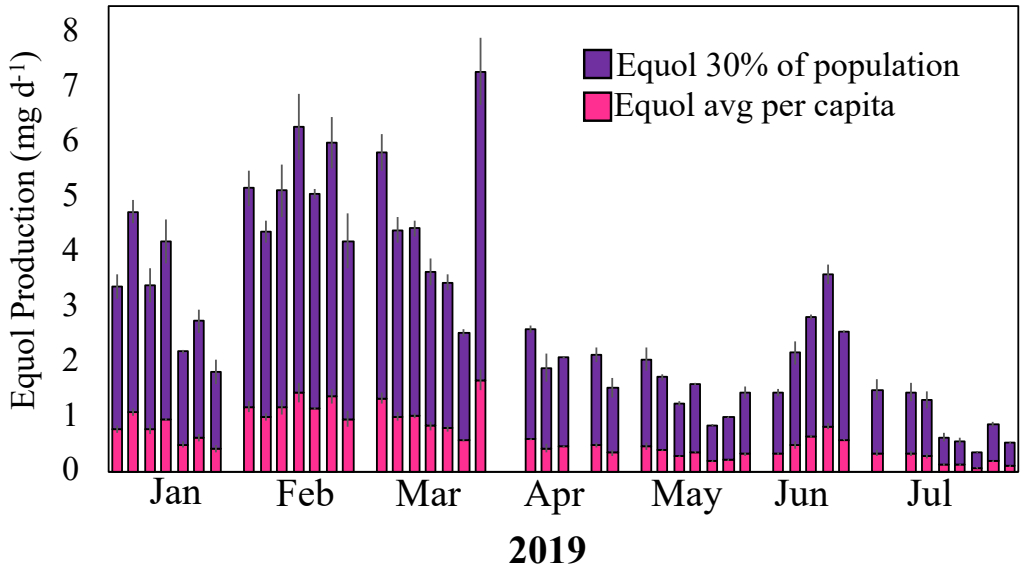

Figure 8. Equol production in mg d^{-1} for 30% of the population in this study and average per capita for entire study. Results here demonstrate a possible solution for reflecting measurements for equol production using WBE given ~30% of a Western population can produce equol. Shown in pink is the average per capita production measurement using the entire population of this study which may underestimate the amount of equol produced in a given population. Shown in purple is the equol production corrected to reflect one-third (30%) of the measured population in this particular study. Measurements for equol began in January of 2019 and lasted through July 2019. Error bars represent the standard deviations for replicate samples.

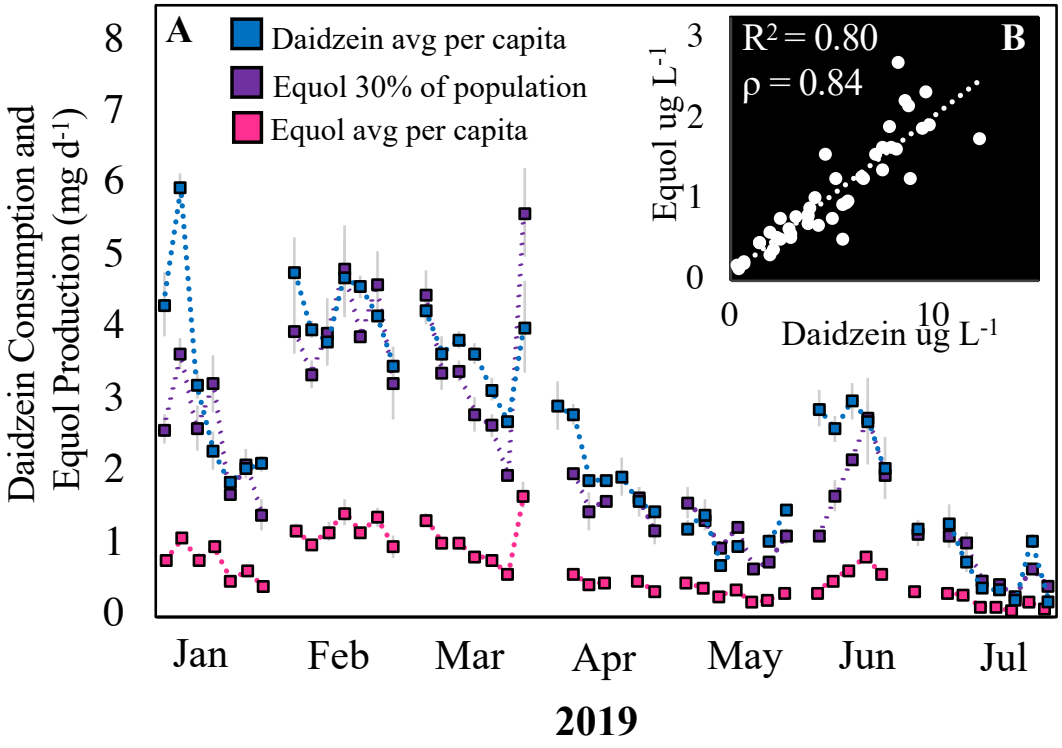

Figure 9. Understanding daidzein and equol interactions at population-level using WBE. **(A)** Daidzein (parent) consumption, shown in blue, and equol (metabolite) production, shown in purple as 30% of population and in pink as average per capita for full population under study. Measurements took place from January through July 2019 (7 months) and shown in **mg d^{-1} per capita**. Trends in daidzein and equol mimic trends shown for genistein, daidzein, and enterolactone. Equol production corrected for 30% of the population may reflect more accurate measurements rather than for average per capita, which appears to underestimate measurements. **(B)** XY Scatter plot of measured concentration of daidzein and equol (μg L^{-1}) in upper right corner displaying strong linear relationship (R^2 = 0.80) and strong positive association (ρ = 0.84) tested using Spearman's Rank nonparametric test.

For all compounds examined, there was a unique spike in March either on or immediately following the holiday, St. Patrick's Day, in both 2018 and 2019 (March 17th, 2018 and March 18th, 2019), with the exception of enterolactone in the 2018 measurement, which could be explained by the extended half-life compared to the other measured analytes. This could be from increased consumption of traditional Irish foods in celebration of the holiday, such as colcannon or corned beef and cabbage, which all contain varied amounts of lignan and isoflavones. However, more likely, this could be due to an increase in alcohol consumption, particularly beer, which is known to contain certain types of phytoestrogens, including lignans [145]. Interestingly, in 2019, equol levels exceeded those of its precursor, daidzein (**Figure 9A**). Given the nature of the holiday, alcohol consumption could be a potentially acceptable explanation, warranting further investigation. In a study comparing equol-producers and non-equol-producers, those who reported higher alcohol consumption were more likely to be stronger equol producers, however, the mechanism behind this is still underdeveloped [146]. Alcohol consumption has been indicated to increase over this particular holiday, and was measured at elevated levels using WBE in March of 2018 in this general location [124], and repeated in 2019 during this study (**Figure 19**). This could explain the elevated equol levels measured in March of 2019 (**Figure 8**).

Comparisons with current estimates for phytoestrogen consumption informed by conventional methods for nutritional assessment with measurements in this study were performed, starting with a cost analysis which resulted in a per person cost of <$1 for a WBE study (10,000 population) with existing startup equipment. This number increases to just over $2 per person if startup equipment is needed, such as an LC-MS/MS

(Appendix A: **Table 16**). This is considerably less than average cost estimates for traditional nutritional assessment, which could cost >$15 per person depending on the instrument used [4, 103].

4.3.1 Limitations

Comparisons between the measured signal in wastewater for soy isoflavones and lignans and reported estimates of consumption based on individual studies and self-reported survey data were limited largely due to differences in resolution between the two sets of data, resulting in varied demographic distributions. Ranges are often provided as a result from large-scale survey methods for overall phytoestrogen consumption in a western society, however, demographics at a national scale will vary significantly than at the sub-city scale. Variations in demographics will also result in interindividual variability in consumption patterns, for example, an individual who is not following a vegetarian diet will likely receive most of their phytoestrogen intake through lignans, however, flaxseed is the densest of lignan sources, and is not a food commonly consumed in a western-type of diet. While this poses as a limitation for comparing to international estimates, the purpose of conducting WBE from within the sewer catchment was to provide contextual information of the study of interest, thereby allowing for targeted and relevant public health interventions, if warranted. Thus, this study highlights the importance of more alternative approaches to population health as a function of demographics.

Additionally, loss or reduction of analyte signal during travel time from within the sewer pipe were not accounted for in this study. Stability studies specific to these compounds have reported that they are relatively stable; noting analyte degradation only

at extreme temperatures (70-90℃) that are not typical of a wastewater matrix (30-35℃) [147]. Additionally, in-sewer travel times in this study did not exceed 50 minutes (20 minutes on average); this is significantly shorter than most WBE studies, which can serve to enhance the measurements reported herein.

4.4 Conclusions

This study investigated several aspects of assessing population-level dynamics of dietary behavior by integrating a multi-omics approach to wastewater-based epidemiology. While previous studies have reported the potential for dietary assessment through WBE, the true feasibility had yet to be tested from multiple perspectives and on a long-term scale to determine unique, contextually-relevant trends in the study sample. Thus, this study offers several lessons learned to inform future studies, as it is the first to report *(i)* long-term (two-year) consumption patterns of a plant-based diet at the neighborhood level, *(ii)* observe an association between wastewater-informed levels of phytoestrogens, as well as human health-related biotransformative products and the composition of microbial communities presumably associated with human gut microbiomes, and *(iii)* the introduction of equol, a microbial metabolite of daidzein, in order to validate the use of daidzein as a dietary biomarker under certain conditions while also refining wastewater-derived consumption estimates. Further, the use of equol as a clinical intervention for certain types of hormone-related conditions is intriguing to explore in future studies, as a population-level health assessment informed by WBE could serve to provide a comprehensive understanding of the benefits of equol from a holistic model. The data reported here suggest that a comprehensive approach to

understanding dietary behavior at population-scale is essential in order to provide the most relevant source of data available for any given community. Given the potential for WBE to provide actionable and useful information for public health assessment and intervention, this study serves as a foundation for future implementation.

CHAPTER 5

UNRESTRICTED, ONLINE SHARING OF HIGH-FREQUENCY, HIGH-RESOLUTION DATA ON SARS-COV-2 IN WASTEWATER TO INFORM THE COVID-19 PUBLIC

Introduction

In Chapter 4, I employed WBE to better understand dietary behavior at the neighborhood-level by implementing a two-year monitoring study of phytoestrogens genistein, daidzein, enterolactone, and equol. An isolated sewage catchment comprised predominantly of a residential population was chosen to carry out this novel pilot-level study. Statistically significant seasonal changes in consumption (mg d^{-1} per capita; fall and winter; spring and summer) were observed, and average consumption rates for isoflavones (genistein and daidzein) were in agreement with current estimates for U.S. average consumption. These findings indicate the measured values in wastewater are reflective of consumption patterns, however, food consumption and human metabolism introduce certain complexities that should be considered for WBE, such as the human gut microbiome. Due to the interindividual variability of the human gut microbiome, these products may produce different results across communities if not appropriately considered during analysis. Further, certain products may also offer unique health benefits than their precursors, and thus warrant further investigation for future public health applications. Thus, a microbial composition analysis was performed to test if taxa exclusive to the human gut responsible for biotransformation of consumed foods could be identified in community sewage. Further considerations for dietary analysis by WBE was discussed for future study.

In Chapter 5, I explored broader applications of WBE to inform the public health response to the COVID-19 global pandemic that manifested in the United States in early 2020. Challenges experienced early on using traditional methodologies for reporting viral presence highlighted the need for alternative surveillance strategies, such as WBE. This study adopted a high-resolution, high-frequency approach to sample collection, where three samples were collected from eleven catchments per week that encompass the entire City of Tempe, AZ. Resultant wastewater data (viral genome copies L^{-1}) were reported weekly to city personnel and subsequently shared on a publicly available online dashboard. This was employed to test the ability for WBE to identify potential hotspots from within the city as well as to provide an early-warning signal compared to conventional reports. Results suggest that wastewater surveillance accomplished these goals, demonstrating the importance to include WBE for public health surveillance strategies. Lessons learned from employing this highly collaborative effort were discussed, with suggestions for future investigation.

5.1 Wastewater-based epidemiology to monitor for SARS-CoV-2

Triggered by the SARS-CoV-2 pandemic, the use of wastewater-based epidemiology (WBE) as a potentially powerful, rapid, and inexpensive tool to inform public health decision-making has seen a remarkable increase globally. For decades, WBE has been exercised to track chemical and biological threats, with numerous studies underscoring its efficacy and usefulness for understanding and managing community health [148-155]. At the onset of the SARS-CoV-2 pandemic, significant delays in conventional and individualized clinical testing, due in part to an overwhelmed healthcare system and

resource limitations [156], positioned WBE as a promising supplemental tool for assessing SARS-CoV-2 spread at the population-level, a strategy that soon was adopted more broadly [157-160]. Early data showed SARS-CoV-2 levels in wastewater and sludge as a concomitant or early indicator of clinical confirmed infections, disease and mortality in a community [161, 162].

The City of Tempe, Arizona, residential population ~200,000, had been an early adopter of WBE for the purpose of tracking opioid consumption, which began in May of 2018 and led to the launch of a fully interactive, public-facing, open access WBE dashboard in February of 2019 [163]. In a municipal-academic partnership, Tempe and Arizona State University (ASU) participated in sharing of monthly wastewater samples, subsequent analysis, and to joint reporting of use-trends of opioids within the community monthly by displaying the obtained collaborative results for oxycodone, codeine, heroin, and fentanyl (and metabolites; $\mu g\ d^{-1}$ per 1,000 people) in five urban sewersheds [164]. The City also had established a routine for data analysis and public health response by integrating Tempe Fire Medical Rescue, Human Services (e.g., CARE 7 crisis intervention organization), and others into a workgroup that relied on WBE data as an important and innovative source of information to guide resource deployment by community need (**Figure 10**).

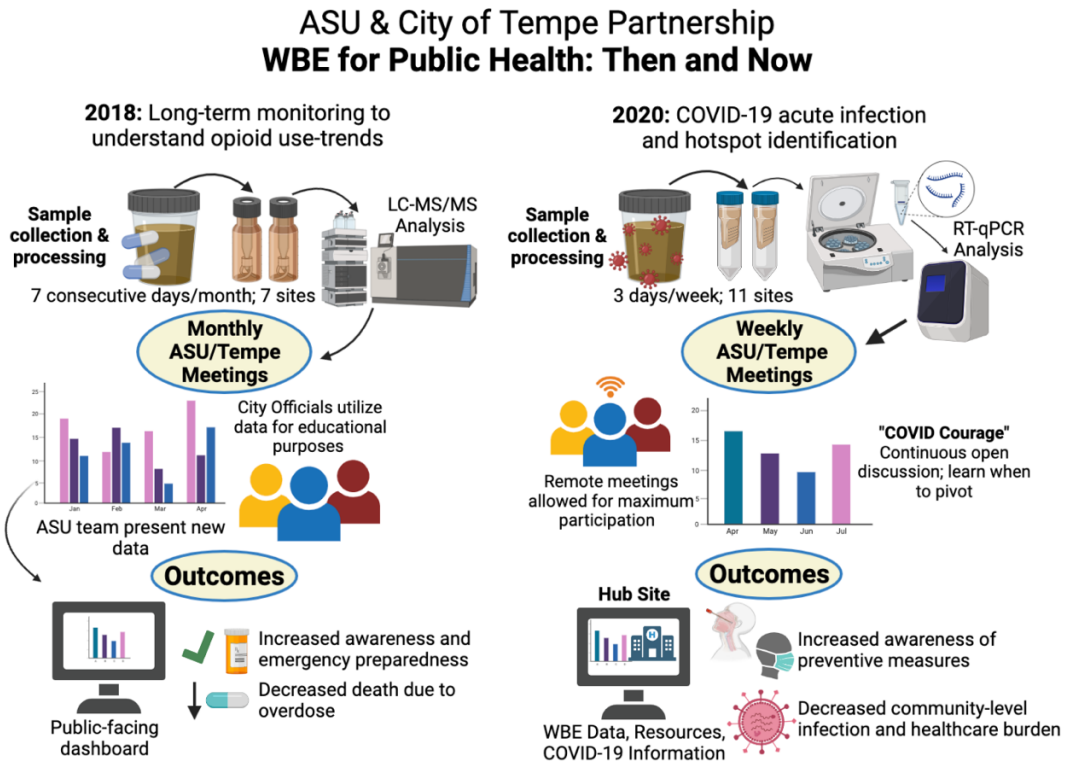

Figure 10. Schematic of the ASU/Tempe partnership, demonstrating how the existing wastewater monitoring network for opioid use (established in 2018) enabled a rapid transition to monitoring SARS-CoV-2 during the COVID-19 global pandemic (2020) with work products including the world's first WBE-informed public interactive online dashboards to combat the opioid and COVID-19 epidemics through a data-driven targeted public health response. Created with BioRender.com.

With this existing framework in place, Tempe and ASU were in a unique position at the start of the SARS-CoV-2 pandemic to quickly transition into WBE surveillance of SARS-CoV-2. Coincidentally, Tempe and the ASU community also had one of the first early diagnoses of a positive SARS-CoV-2 patient (26 January 2020)[165]. As the City of Tempe and ASU quickly transitioned into molecular-based monitoring, the immediate goal was to use previously established expertise in sampling, infrastructure access, and WBE-framed public health response to begin quantitative assessments of SARS-CoV-2

levels in wastewater. The ultimate objective was to identify hotspots of infection early and implement interventions including education, outreach, and targeted clinical testing to limit the spread of the virus within the Greater Tempe community. The local health department shared data from clinical testing of individuals only at the zip code level, [166] a policy intended to protect small communities and personal identifiable information, which potentially limited stakeholders' ability to respond to local virus clusters. Unique to the US, zip codes are a series of 5 numbers created by the US postal service to delineate small geographical areas within counties to improve mail service, and are used extensively by local and state agencies, including public health departments [167]. The 5-digit, Tempe, AZ zip codes involved in this study are 85281, 85282, 85283, and 85284, and will be referred to here as ZC-1, ZC-2, ZC-3, and ZC-4. The pre-existing, neighborhood-level wastewater monitoring network offered an opportunity to test the potential of WBE to serve as an early warning system that may reveal virus presence and spread prior to clinical case data reported from testing of individuals [168, 169]. Thus, important goals of the work were (i) to compare WBE data to newly reported clinical cases of SARS-CoV-2, related hospitalizations, and associated deaths at a high temporal and geospatial resolution (i.e., county, city, zip code, and neighborhood levels), and (ii) to determine whether the concurrent pandemic monitoring by WBE produced data and information not available or obvious from clinical testing.

5.2 Materials and methods

5.2.1 Study Location

This study was conducted within the City of Tempe, Arizona and the Town of Guadalupe, Arizona, (i.e., Greater Tempe), with an estimated residential population of approximately 200,000, and home to Arizona State University, one of the largest public universities in the US. The City was divided into nine sewer catchment areas predetermined by the City for regulatory compliance monitoring purposes and for ease of access in the scale-up of this project. Two additional sampling locations were also included to isolate Tempe-only SARS-CoV-2 signals.

5.2.2 Sample Collection

Flow- and time-weighted 24-h composite samples of untreated wastewater were collected at each sampling location within the wastewater collection system for three days each week (Tuesday, Thursday, Saturday), beginning April 2020 (Catchment 7 and Tempe St. Luke's Hospital were added in July 2020). Samples were collected either with an Avalanche refrigerated sampler or a portable sampler (Teledyne ISCO, Lincoln, NE) using a mixture of wet and dry ice for cooling. Units were equipped with 9 mm inner diameter (ID) silicon tubing on the pump head, and silicon PTFE lined tubing of the same diameter. Flow was monitored by an ISCO LaserFlow meter (Teledyne ISCO, Lincoln, NE), located within a nearby manhole or flow estimated based on historic data. Composite samples were collected in acid-washed bottles and transferred to high-density polyethylene bottles that were quickly placed on ice in coolers for transport. Samples were processed immediately same-day to minimize degradation losses.

5.2.3 Sample processing and analysis

Raw wastewater samples were analyzed for SARS-CoV-2 RNA following sequential steps of filtration, concentration, nucleic acid extraction, and reverse

transcriptase quantitative polymerase chain reaction (RT-qPCR) analysis. Approximately 150 mL aliquots of raw wastewater samples were filtered through a sterile 0.45 μm polyether sulfone (PES) membrane filter unit (Fisher Scientific, Lenexa KS) by vacuum for removal of large debris. The filtrate was then loaded onto two Amicon® Ultra 15 centrifugal filters (Millipore Sigma, Burlington, MA) with a 10,000 molecular weight cutoff and centrifuged at ~4,000 RPM for 15 to 20 minutes for five sequential intervals with an Eppendorf 5810R swing bucket refrigerated centrifuge (Eppendorf, Enfield, CT). The final concentrate was combined into 1.5 mL conical microcentrifuge tubes, and 200 μL was processed using an RNeasy mini extraction kit from Qiagen (Germantown, MD), modified for use with this specific matrix following an animal cells protocol. The extracted RNA (50 μL) was stored at -80°C until quantification by RT-qPCR using SuperScript III Platinum One-Step qRT-PCR Kit (Invitrogen, Carlsbad, CA). The Charité/Berlin (World Health Organization) designed primers and probe for the E (envelope) SARS-CoV-2 gene target were purchased from Integrated DNA Technologies (Coralville, IA) [170, 171]. For quality assurance and quality control, deionized water was used for whole process negative extraction controls for each sample batch and RNAase/DNAase free Ultrapure™ water (Invitrogen,Waltham, MA) was used as molecular negative template control along with a SARS-CoV-2 positive control in every RT-qPCR plate. The positive control was created by *in vitro* transcription using linearized plasmids, with subsequent sequencing to determine validity. Later this transitioned to a commercially available synthetic full genome target provided by Twist Bioscience (San Francisco, CA). Standard curves ranged from 10^0 to 10^6, with a detection limit cutoff for quantification based on the standard curve (~100 copies uL^{-1}). Triplicate

standard curves were analyzed for each new batch of assay reagents and were used to quantify samples. Quantification was performed using an Applied Biosystems QuantStudio™ 3 Real-Time PCR System with the QuantStudio Design and Analysis Software 1.2 from Thermo Scientific (Waltham, MA). Method details are previously published [172].

5.2.4 Population estimates

Resident populations for each sewershed were estimated using 2010 census block group data. Employment estimates were obtained from the Maricopa Association of Governments (MAG) 2019 employment data and included the following classifications: employees living outside of Tempe (non-resident, employed) and Tempe residents (resident, employed). To correct for changes in employment numbers during lockdown events (commercial closures) and telecommuting activities, we used available MAG average weekday traffic volume (compared to normal conditions) in Maricopa County. This percentage was used to correct the non-resident (Tempe employed) employment numbers. Student population estimates were obtained from publicly-available campus resident data and estimates using changes in wastewater flow volume (Appendix B: **Table 17**).

5.2.5 Clinical data

Newly detected clinical cases by zip code within the City of Tempe, Arizona were reported daily by the Arizona Department of Health Services. The City of Tempe began extracting and archiving these data on 23 May 2020. Prior to this, daily case data are not available (data are in aggregate as total cases from the start of the pandemic). Maricopa County-level new positive cases, COVID-related hospitalizations, deaths and long-term

care facility deaths per day are publicly available and were collected from the Maricopa County Epidemic Curve Dashboard [166].

5.2.6 Data and statistical analysis

Measured concentrations in each sewer catchment were transformed to viral load (VL) per day (genome copies d^{-1}) using the following equation:

$$Viral\ load\ (genome\ copies\ d^{-1}) = C_x\ x\ Q_x \qquad \text{Eq. 7}$$

where C_x (genome copies L^{-1}) is the measured concentration in a given catchment, Q_x is the total daily volumetric flow rate (L d^{-1}). In cases where one sewer catchment flowed into another, viral loads were subtracted to isolate the individual catchments.

Statistical assessments were conducted in MATLAB R2021a (MathWorks, Natick, MA). Root mean square error (RMSE) was used to calculate the offset between different compared data categories using the following equation:

$$RMSE = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad \text{Eq. 8}$$

where n is the number of observations, x_i the viral loadings of SARS-CoV-2 in wastewater, and y_i either the newly detected clinical cases, COVID-related hospitalizations, or COVID-related deaths. Data were assessed between 1 April 2020 and 31 March 2021, using individual waves of infection corresponding to up to three events peaking in June/July 2020, August 2020, December/January 2020-21. Data were shifted from 0 to 20 days in both directions for each of the comparisons. The data resolution between clinical cases and wastewater testing were different (daily vs. 3x per week), so clinical results that did not have a corresponding wastewater data point were omitted from the assessment, post shift.

5.3 Results

5.3.1 Neighborhood-level sampling

At the onset of the pandemic, our team had divided the Greater Tempe area into five sewer catchments (Areas 1-5; **Figure 11**), including two additional, non-published locations that received wastewater from adjacent municipalities, which were necessary to determine the Tempe-associated sewage signal where wastewater was comingled. The neighborhood-level sampling methodology was synchronized with reoccurring compliance monitoring of the Sub-Regional Operating Group (SROG), a cohort of five municipalities including Phoenix, Tempe, Mesa, Glendale, and Scottsdale, that jointly own and operate the 91st Avenue wastewater treatment plant (WWTP) in Phoenix, Arizona. The predefined sampling strategy consisted of 7-consecutive days of sample collection each month, across variable weeks from permanent, sub-surface sampling stations. While this sampling strategy was sufficient for long-term, opioid-related monitoring, tracking of SARS-CoV-2 levels required an increased temporal resolution. Accordingly, we adopted a high-frequency sampling approach consisting of weekly collection on Tuesday, Thursday, and Saturday, in addition to the SROG sampling events. To improve spatial resolution, additional sampling locations were also identified based on ease of collection (Area 6), while three other permanent locations needed slight infrastructure modifications (Area 7) and/or approvals prior to onboarding, including the Town of Guadalupe with strong representation by Native American and Hispanic residents (**Figure 11B**), and Tempe St. Luke's Hospital (not displayed on dashboard). Permanent sampling locations outside of the Tempe jurisdiction (necessary for eliminating non-Tempe SARS-CoV-2 signals) were available only during the previously

referenced week of compliance monitoring. City of Tempe personnel began sampling from maintenance holes (also known as manholes) immediately downstream of these locations, within the City's jurisdiction, during the three weeks each month when regular compliance sampling was not performed.

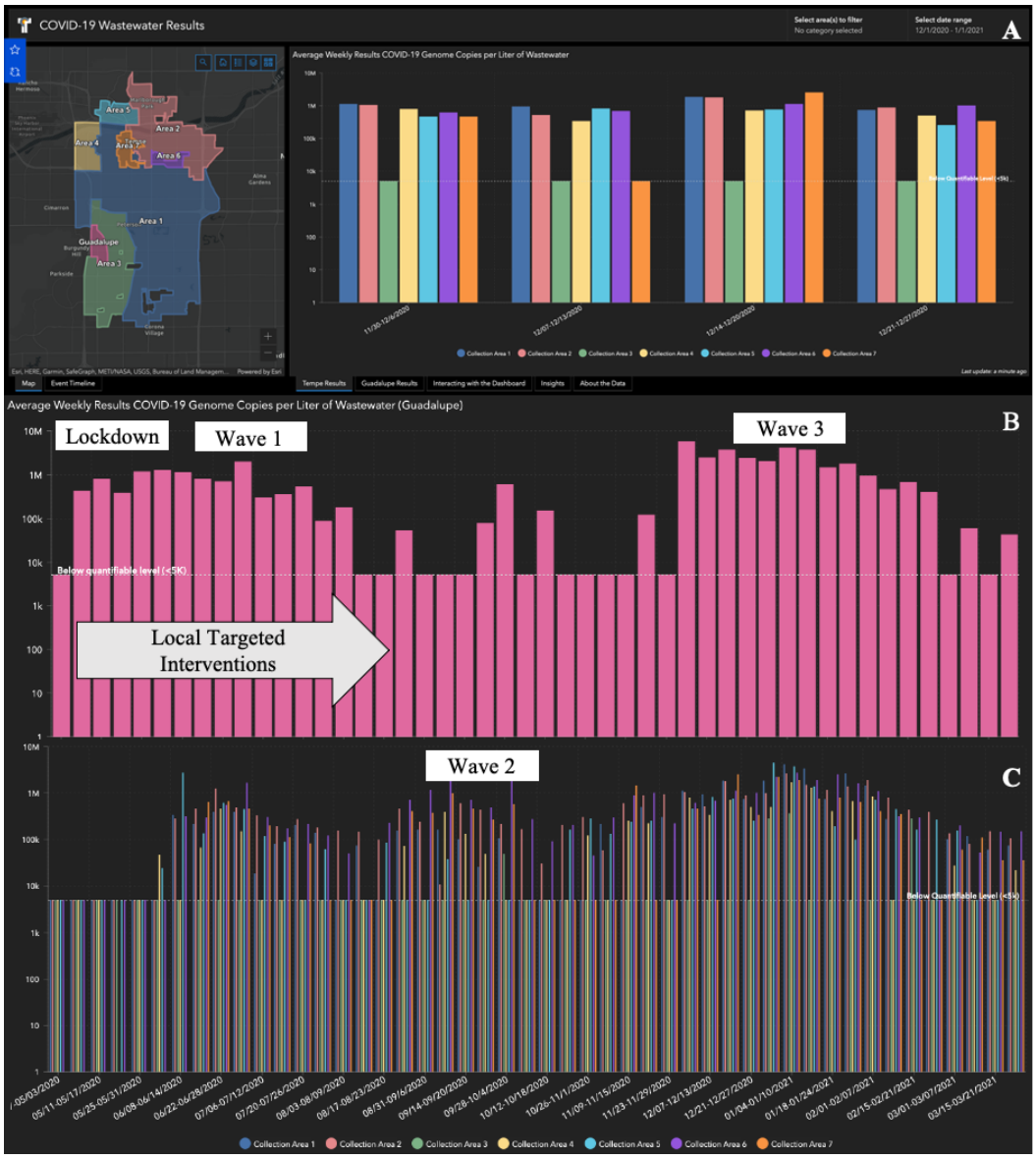

Figure 11. Publicly accessible, interactive WBE online dashboard showcasing **(A)** a Map of Tempe, Arizona divided into wastewater catchments (Areas 1-7 and Town of Guadalupe) alongside one month of SARS-CoV-2 levels in wastewater determined by targeting the E gene, with additional data being displayed for **(B)** The Town of Guadalupe (May 2020 through March 2020); with local targeted interventions implemented from May 2020 through August 2020 and **(C)** The City of Tempe (May 2020 through March 2020), showing the impact of the initial lockdown, and subsequent waves of infection over time (viewable at covid19.tempe.gov).

5.3.2 Determination of SARS-CoV-2 levels in wastewater

Between 1 April 2020 and 31 March 2021, a total of 1,556 samples were collected across the Greater Tempe area. Each 24-h composite sample collected represented either

a flow-weighted or time-averaged sample (15 min. collection intervals) captured by a high-frequency automated sampler. The number of samples collected per catchment during the study varied from a high of 155 in Area 1 to lows of 103 (Area 7) and 101 (Hospital), with observed differences resulting from occasional sampler malfunctioning and staggered onboarding of additional sampling locations. The total number of SARS-CoV-2 detects per catchment throughout the study was on average 66 ± 36, with a minimum of 4 (Area 3) to a maximum of 116 (Area 6). On average, when detected, the coronavirus concentration was 617,000 ± 2.075 million (M) gene copies L^{-1} (median of 251,450 gene copies L^{-1}) indicative of great fluctuations in virus levels over time. Detailed concentration information is provided in **Figure 12**.

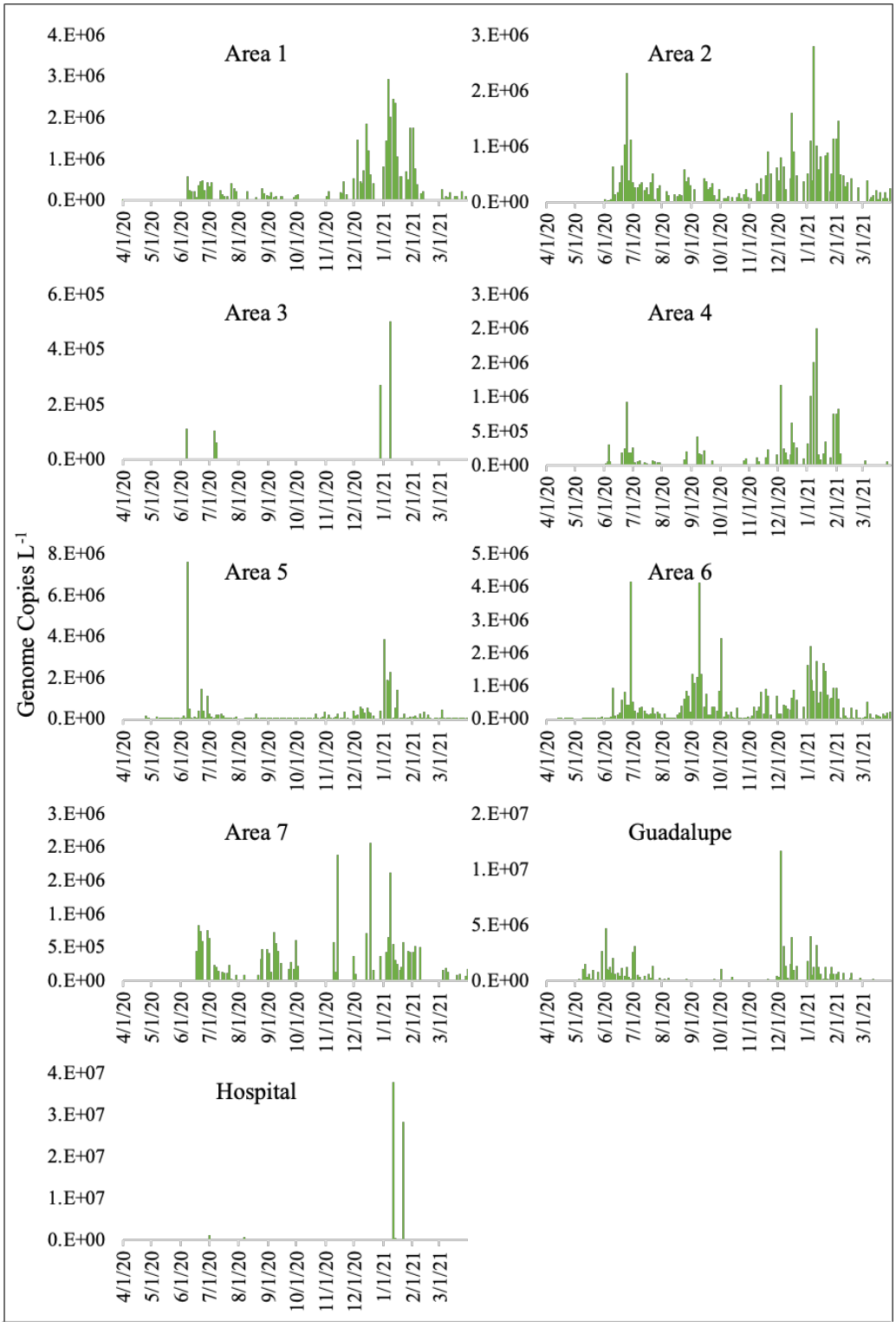

Figure 12. Measured concentration in target catchments (Areas 1-7; Guadalupe; Hospital) from 1 April 2020 – 31 March 2021.

The SARS-CoV-2 viral load was calculated for each sample at a collection point using wastewater flow data provided by Tempe (Appendix B: **Figure 18**). Flow rates in catchment Areas 1-7 had data recorded at 2-min intervals in real-time using permanent laser flow meters, while the Town of Guadalupe and the Tempe St. Luke's Hospital had only historical flow data available. Flow varied from a maximum in Area 1 of 54.5 ± 6.6 million L day^{-1} (MLD) to a minimum of 0.106 MLD (historical estimate) for the hospital location; average flow rates across all catchments were 15.1 ± 21.3 MLD.

At select collection sites, the corresponding wastewater sample was representative of multiple collection catchment areas due to the comingling of wastewater in the collection system (Appendix B: **Figure 19**); this occurred in Areas 1-3. To isolate individual catchments and provide a catchment-specific viral load, a mass balance was performed. Resultant viral loads in each sewer catchment ranged from 6×10^{10} - 1×10^{13} genome copies d^{-1} (**Figure 13**). The distributions in viral load varied between each location, with Areas 1, 2, and 6 showing smoother distributions in viral load over time, while others showed more isolated single-day spikes in activity. Most locations showed two waves in viral levels occurring in June 2020 and December/January 20-2021. However, in catchments close to ASU, an additional unique wave of viral load was visible (Areas 6 and 7) in August 2020 (**Figure 13**).

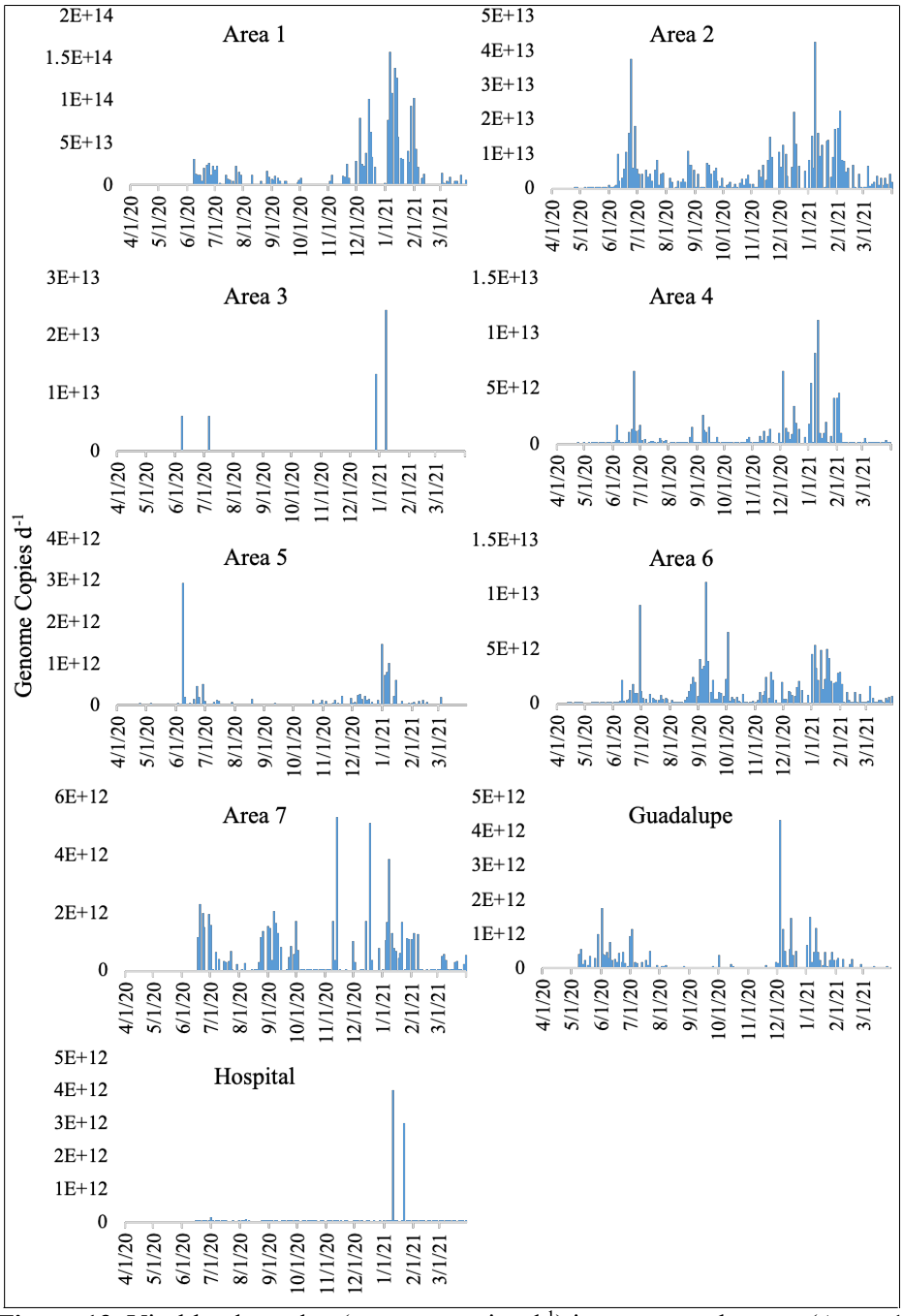

Figure 13. Viral load per day (genome copies d^{-1}) in target catchments (Areas 1-7; Guadalupe; Hospital) from 1 April 2020 - 31 March 2021.

Understanding the number of people contributing to wastewater in any given catchment is critical when working with data generated from within the collection system. Estimated Tempe subpopulations ranged from a low of 8,114 ± 848 in Area 5 to a high of 132,082 ± 7,374 in Area 1, the largest geographic catchment area (Appendix B: **Table 17**). Variability in Tempe data was a function of the total numbers of residents, employed individuals, and the number of students in the contributing area. The population of the Town of Guadalupe (6,500) was determined using US census data [173]. The hospital location was omitted from this population analysis since the number of individuals working or serving as patients was unknown.

5.3.3 Data usage for public health

The result of these efforts ultimately culminated in eight SARS-CoV-2 collection locations viewable online by the public (Areas 1-7 of Tempe and Town of Guadalupe) on an interactive dashboard that went live the first week of May 2020 (**Figure 11**). The dashboard displays each catchment area overlain on a street-level city map so users can geospatially identify contributing locations in the catchment. In response to a request of the impacted communities, the Town of Guadalupe is displayed on a separate tab of the dashboard. Data are shown as the logarithm of genome copies L^{-1} and are presented as a weekly average consisting of the Tuesday, Thursday, and Saturday collected samples. Since the sewage collection system in Tempe separates stormwater from municipal wastewater and the study site is in an arid climate, the use of concentration was permissible. Users have the ability to control spatial and temporal parameters to their preference; and text and infographics accompany these data, which explain WBE basics, how to properly interpret the data, and how data are created and used by the City.

Additionally, the SARS-CoV-2 wastewater dashboard is nested in a Community COVID-19 Health Site that contains information provided by the Centers for Disease Control and Prevention (symptoms, prevention, exposure-response), City demographic information, and positive clinical cases reported by zip code. To ensure congruency in data interpretation between Tempe and ASU, the frequency of joint meetings that began in 2018 increased from monthly to weekly beginning April, 2020 or biweekly (December, 2020 and on) to discuss wastewater-derived data, newly reported cases in the community, sampling logistics, and targeted mitigation strategies in communities in response to the collected data when applicable. Although sewage temperature, travel time, and storage are known to influence the stability and signal strength of labile wastewater-borne biomarkers such as virus RNA [174], a lack of temperature-specific attenuation rates of SARS-CoV-2 did not allow for a confident normalization of data during the time of study. However, the short residence times within the various neighborhood-level sewersheds in Tempe and use of refrigerated samplers decreased the impact of these variables on collected samples.

Data generated from the SARS-CoV-2 wastewater monitoring revealed consistently elevated virus measurements in the Town of Guadalupe during the initial lockdown in May to early June 2020, a marked difference compared to the other catchments in the area. A targeted approach of implementing face-mask mandates and community education in town halls, adding clinical testing sites, in addition to public sharing of wastewater data occurred from April through August, 2020, resulting in the decline of both viral loading in wastewater and newly reported SARS-CoV-2 positive cases. The Town of Guadalupe's wastewater comingles with Tempe wastewater (and an

additional external municipality) and is collected as a composite at Area 3. However, Area 3 only had four detects during the entire year-long sampling campaign, implying that the elevated SARS-CoV-2 signal originating in the Town of Guadalupe was attenuated beyond detectability at the Area 3 sampling location, and was visible only through high-resolution monitoring. Whereas clinical testing in theory may have rendered the Guadalupe infection hotspot visible, testing was lacking in the area, and newly reported case counts from traditional clinical tracking by the county were aggregated at the zip code level with that of Tempe, thereby obscuring what may have been happening within the community.

5.4 Discussion

We employed WBE to monitor SARS-CoV-2 in the Greater Tempe, Arizona, Southwestern United States by implementing a unique, high-frequency, and neighborhood-level sampling approach in conjunction with immediate, open access data sharing with the public. The present work illustrates how an established WBE monitoring network can be adopted quickly to shift from one public health priority to another, as done here by switching from opioid targets to SARS-CoV-2. Measured values in this study were in line with those reported from other wastewater monitoring studies; [162, 175, 176] however, the maximum concentration of 37.6 M gene copies L^{-1}, is among the highest recorded measured values to date. This measurement occurred at the hospital location, which had an active COVID-19 ward at the time of collection on 11 January 2021 during peak pandemic conditions (to date). Higher relative standard deviations (RSD) in measured SARS-CoV-2 values for a given week (n=3 observations per week),

occurred in locations with a higher proportion of commercial businesses, including Areas 4 and 5 (RSD 83, 93%) as compared to those with largely residential catchments (Areas 1 [58%] and Area 2 [65%]). This may explain the relatively smoother trends over time in Areas 1 and 2, as compared to those with higher transient populations, which showed isolated single-day spikes in viral presence. This is plausible given non-ambulatory individuals likely would have stayed at home throughout the duration of their illness. These results suggest that a high-frequency sample collection approach should be considered in catchments with a higher proportion of transient populations, which may be susceptible to greater variability in virus occurrence from day to day.

Estimating population size by study area was challenging due to the unique nature of collecting wastewater from within the sewer infrastructure rather than by determining the population by counting the residents of local buildings served by a wastewater treatment plant as performed in traditional WBE studies [59, 177]. As a net importer of people to the City for work, it was important not only to quantify the residents but also the non-resident employed and transient student populations, a task accomplished by using Maricopa Association of Governments (MAG) and on-campus student resident data provided by ASU. MAG data needed to be corrected for lockdown activities which closed businesses, for which we used Arizona department of transportation arterial traffic flow data (e.g., 40% decrease in arterial traffic equated to 40% decreased in employment populations). Due to the bulk of student classes moving online, using only campus resident data (based on student housing contracts which were updated monthly) was appropriate to assess temporal changes in student populations. These numbers did not account for changes in resident population during holiday travel or off-campus housing

locations; thus, overall percentage changes in wastewater flow were also used to estimate population size changes. For instance, wastewater flow from Area 6 increased by 20% and was sustained throughout the academic year; therefore, the population in that area was assumed to increase proportionally. This increase in total flow in Area 6 also coincided with increases in viral load, suggesting that infected students were moving back into Tempe for the start of the academic year. Looking ahead to Fall 2021 when classes are expected to resume in-person, quantifying that transient population will become more important for sewersheds impacted by students. We have therefore begun testing the utility of campus Wi-Fi data to better estimate population size as students and faculty return to pre-pandemic campus activities [157].

The measured viral loads per day of SARS-CoV-2 within each catchment area in Tempe were aggregated and partitioned to their respective zip codes (ZC-1 through 4) according to their estimated percent contribution (Appendix B: **Figure 20**). Wastewater-derived SARS-CoV-2 peaks in activity correlated with newly detected clinical cases per day in three distinct waves of activity: June 2020, August 2020, and December/January 2020-21. ZC-1, home to ASU, was the only zip code that showed viral increases in August 2020. Contributions to viral load within a given community by university students, however, is not an event isolated to Tempe [178-180]. Comparisons between spikes of coronavirus levels in wastewater and clinical case data showed that peaks in wastewater preceded positive clinical cases by 7, 6, 11, and 10 days for ZC-1 through 4 (average of 8.5 ± 2.1 days), during the first wave of the pandemic, and again during the isolated university-associated wave, this time by 6 days (**Figure 14**). These results align with preliminary assessments of wastewater and clinical case data that suggested

monitoring wastewater provided an early-warning capacity ranging between 2 and 21 days [162, 175, 181]. Tempe aggregated viral loads were also compared to Maricopa County Public Health data (**Figure 15).** Results again showed peaks in wastewater measurements preceded new clinically reported cases, SARS-CoV-2-related hospitalizations, and SARS-CoV-2-related deaths by 2, 16, and 18 days during the first wave of the pandemic. These results align with prior work demonstrating that wastewater can serve as an early indicator of future clinical case load, morbidity, and mortality.

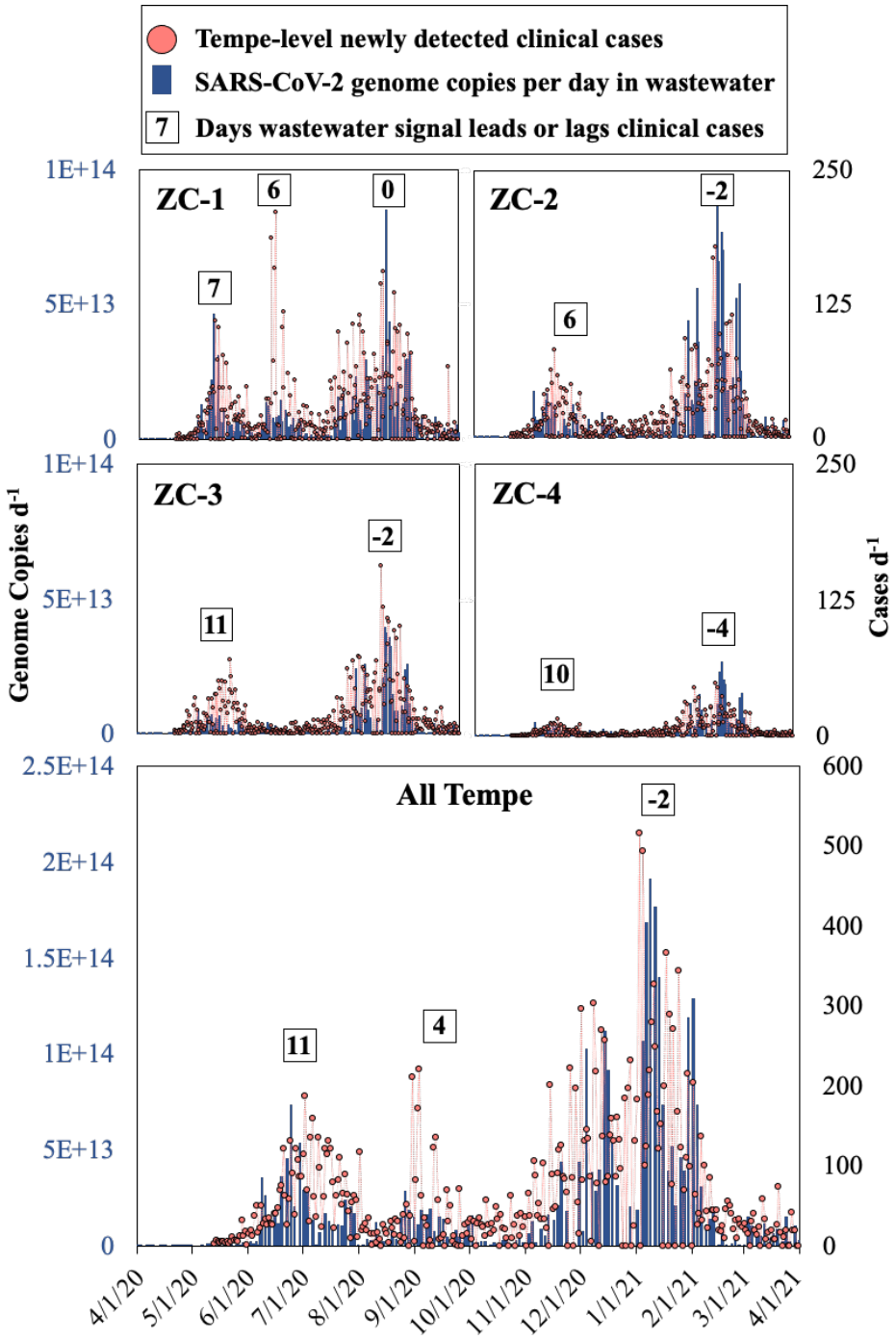

Figure 14. SARS-CoV-2 genome copies per day in the four zip codes (ZC-1 through 4) of Tempe, AZ and in aggregate, overlaid with newly reported clinical SARS-CoV-2 cases. Boxed numbers are the number of days the wastewater signal leads (+) or lags (-) clinical cases, determined by root mean square error (RMSE) analysis.

Interestingly, four months later during the December/January 2020-21 wave of the pandemic, wastewater was no longer a leading indicator in any region in Tempe, AZ. Trends either directly aligned with newly reported clinical cases (ZC-1) or lagged behind clinical case data by 2 days (ZC-2 & 3), 4 days (ZC-4), and 2 days (aggregate). At the county level, wastewater lagged behind clinical results with wastewater peaking 3, 3, and 1 days behind Maricopa reported cases, SARS-CoV-2-related hospitalizations, and deaths, respectively. To our knowledge, no study has reported wastewater as both a leading and a lagging indicator to clinical cases or demonstrated such a transition from lead to lag at the sub-catchment level within the same community and from within the sewer collection system. This consecutive decrease in lead time between wastewater measurements and clinical testing may best be explained by notable differences in the availability and frequency of clinical testing over the course of this study. Qualitative data from Maricopa County shows that access to testing was extremely limited during the early stages of the pandemic [182, 183] but increased dramatically subsequently with the continuous onboarding of commercial, hospital, and university laboratories, largely driven by ASU biomedical screening. Thus, these data strongly suggest that the greatest benefits of WBE are to be verified early on during the detection of disease outbreaks before health care providers can mount a response. Similar benefits may be reaped late into an epidemic, when clinical testing of individuals becomes cost-prohibitive and may appear unproductive when generating mostly negative results.

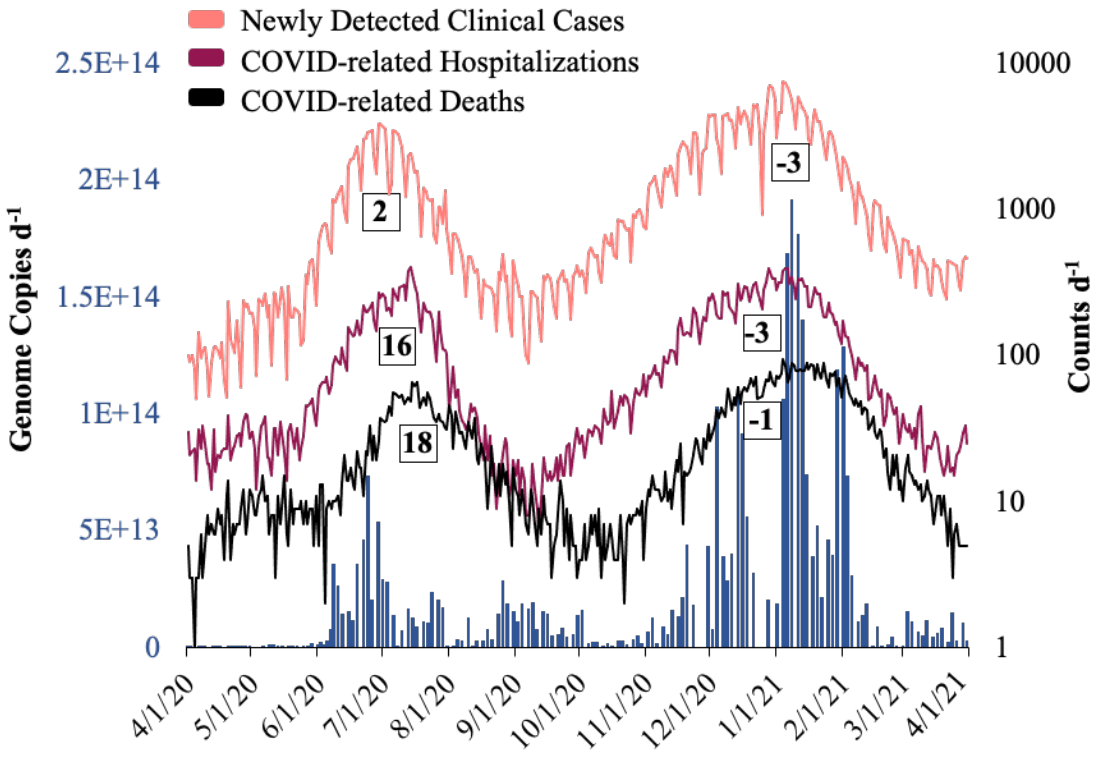

Figure 15. Peaks in SARS-CoV-2 viral load (genome copies d^{-1}) in Tempe, AZ wastewater as compared to Maricopa County, AZ new clinically detected cases, SARS-CoV-2-related hospitalizations, and related deaths. Boxed numbers display days the wastewater signal leads (+) or lags (-) clinical cases.

Results of this work show that the greatest benefits of WBE may be early detection of disease outbreaks in situations where a significant health care response has not yet been mounted, i.e., when clinical testing is still lacking or scarce. Other factors potentially impacting the early-warning characteristics of WBE may include testing fatigue, widespread use of at-home rapid tests, and vaccination campaigns. Thus, the study identified the importance and impact of choosing sampling locations, highlighting that high-resolution or neighborhood-level sampling allowed identification of isolated hotspots of infections that were not visible at larger catchment levels and within zip code level clinical case data. For future investigation, we recommend to conduct sampling, where feasible, in locations where wastewater flow is generated exclusively by the

community of interest. This study demonstrated that subtracting inputs from non-target communities is feasible and can result in unexpected discoveries, as found here with the virus cluster originating from the Town of Guadalupe adjacent to Tempe, AZ. In summary, neighborhood-level sampling comes with increased costs and significant logistic challenges but it also can reveal, as shown here, public health phenomena that are not observable at the city- or zip-code level.

Perhaps most importantly, this study illustrates that a major challenge to neighborhood-level monitoring by WBE is not about assay development in the laboratory, but rather creating the partnerships with city personnel, gaining trust from community members, establishing the sampling network and methodology, understanding which establishments or buildings are contributing to a given collected sample, and how these populations change (e.g., weekdays, weekends, during closures or events). These factors lead to detailing and understanding the primary outcomes of this type of investigation; how data should be protected, shared, and used to inform public health decision-making. Further, the results reported here can inform municipalities interested in adopting and implementing WBE programs to monitor already known and newly emerging public health threats, be they of chemical (e.g., opioids) or biological (e.g., SARS-CoV-2) origin.

This long-term study constitutes a powerful demonstration of employing WBE to collect open access, actionable data that were shown here to directly help inform and shape the public health response. High-throughput monitoring of the E gene of SARS-CoV-2 in Tempe sewage showed WBE to provide an early-warning benefit, particularly in smaller subpopulations, with a temporal and spatial data resolution that exceeded that of clinical healthcare data, which are shared only to a limited degree with local

stakeholders. Use of WBE may also be important for communities with barriers to testing (e.g., lack of access, deficit of testing locations, cost), and testing fear (disbelief), or apathy as the duration of the pandemic continues and vaccination levels rise. Most importantly, WBE performed with a high spatial resolution was demonstrated to increase the ability to identify and localize hotspots of infection, thereby allowing for resources and interventions to be implemented in a targeted and more productive fashion.

Sharing data of significant economic and public health importance in a real time, open access format is often considered controversial, potentially leading to apprehension. However, leading up to, and during the study, the City's commitment to open communication in town halls and open-attendance meetings increased transparency and trust from the community. The actions and public health outcomes achieved with this strategy here certainly appear to have outweighed any potential concerns. As such, the City of Tempe is now exploring the applicability of this methodology to other general markers of community behavior and health status. The lessons reported here may inform other communities interested in adopting this new approach, serving as a foundational framework for integrating WBE into public health monitoring and the design and implementation of intervention strategies.

CHAPTER 6

RESEARCH IMPLICATIONS AND RECOMMENDATION FOR FUTURE WORK

The work presented here highlights the need for alternative approaches to understanding human nutrition, chronic illness, and infectious disease at population-level for informing public health strategies and interventions. As reported here, current strategies are largely dependent on subjective measurements or individualized diagnostics, which could result in generating gaps in data, delayed reporting, or resource exhaustion. Population-level assessments that focus on human health, behavior, and activity, such as wastewater-based epidemiology (WBE), can assist current methodologies by operating in tandem in order to fill in these data gaps, and generate community health information in a timely, inexpensive, and culturally competent manner.

6.1 Nutritional assessment and links to chronic disease

In Chapter 2, an extensive literature analysis was performed with two major objectives: (*i*) identify links between diet and disease, and (*ii*) propose a biomarker suite for population-level dietary analysis using wastewater-based epidemiology (WBE). Due to the success of illicit drug monitoring across populations and cities on a domestic and international level, the application of diet had been proposed as a logical next step for future investigation and experimentation. With the incidence of chronic disease, especially those that are related to nutrition behavior such as obesity and diabetes, there has been a greater interest in attempting to define and quantify these links amongst the scientific community. Traditional methods that remain dominant in the field of nutrition to assess dietary behavior rely on self-reported data by the participant. The use of these

methods is informative, validated, and important for gaining an understanding of dietary patterns and behaviors, however, limitations experienced such as recall bias are known to the field and can serve to create or worsen gaps in data acquisition depending on the primary research outcome or experimental design. Investigation of over 30 papers in this literature review indicated, however, measurable links specifically between red meat and phytoestrogen consumption, with reported odds ratios (OR) as either promoting (red meat) or preventing (phytoestrogen) diseases such as cancer and type 2 diabetes. These results suggest a greater need to further refine nutritional assessment at population-scale in order to operate at a comparable pace as chronic disease incidence. Further it was also noted that chronic consumption of a plant-based diet as well as incorporation of whole-wheat and rye can serve to reduce incidence of obesity as well as prevent development of other types of chronic disease, such as cancer. From this analysis, a biomarker suite of the following was proposed for future investigation of a WBE study: indicators of phytoestrogens (isoflavones genistein and daidzein and lignan metabolite enterolactone), red meat (metabolite 1-methylhistidine), whole wheat and rye (alkylresorcinols 3,5-dihidroxybenzoic acid (DHBA) and 3-(3,5-dihydroxyphenyl)-1-propanoic acid (DHPPA)), and cruciferous vegetables (isothiocyanate *N*-acetyl-*S*-(*N*-allylthiocarbamoyl) cysteine). While these indicators have been identified in a human urine matrix, it is not certain that analyte integrity will remain in a complex matrix such as wastewater. While this had yet to be assessed for the majority of the proposed compounds, investigation from a small catchment within the sewer infrastructure could improve chances of analyte detection by minimizing the potential for compound degradation.

Thus, this literature analysis served as a proof of concept to introduce the potential for nutritional status to be identified and monitored using an objective population-level assessment such as wastewater-based epidemiology. The next steps for this chapter would be to test each of the proposed compounds in a WBE study, performing method development and quality assurance measures to define parameters for reproducible investigation across research groups.

In Chapter 3, a secondary literature analysis was performed to further examine the relationship between dietary ingestion of endocrine disrupting chemicals and breast cancer incidence. It was identified in literature in Chapter 2 that there is a peculiar association between phytoestrogen consumption and prevention of breast cancer due to the ability to mimic estrogen or interact with the receptor. This prompted further investigation of exposure to these chemicals referred to as estrogen-mimicking endocrine disrupting chemicals (EEDs) and breast cancer incidence. As mentioned, breast cancer impacts a significant number of individuals worldwide, with an incidence rate in the United States typically impacting approximately one in eight women per year. Understanding interactions that cause breast cancer has become of great interest given this high incidence rate. A type of breast cancer referred to as Hormone Receptor Positive (HR+) is the dominant type, approximately 70-80% of cases, and occurs due to sensitivity of changes in estrogen and progesterone levels in the body. Additionally, it was noted in literature that approximately one-third of all breast cancer cases are considered to be related to dietary behavior. This brought forth the question of dietary exposure to EEDs and whether they could potentially play a role in breast cancer incidence. It was identified in this literature review that on a regular basis, humans are

exposed to EEDs through diet, predominantly through packaged materials, such as (bisphenol-A; BPA), pesticide residues on fruits and vegetables, notably dichlorodiphenyltrichloroethane (DDT) and atrazine, and phytoestrogens' daidzein and genistein. Interestingly, all have been found to have the ability to interact with the estrogen receptor or mimic mammalian estrogen. Estimated ingestion rates, body burden, and reference doses were all assessed in the recent literature (<5 years) in order to gain an understanding of the degree of exposure, whether acute or chronic, and how this information could be relayed to prevent unnecessary exposure. It was determined that the rate of exposure per day was far less than the estimated reference dose when there would be a potential issue for acute toxicity, however, chronic exposure through constant ingestion of these chemicals in low amounts could raise the risk of bioaccumulation and warrants further study to understand tissue distribution, length of tissue occupation, and how that corresponds to breast cancer incidence. Further study into these areas could help to understand the interactions between EEDs, breast tissue disruption, and cancer incidence. Further, it would be prudent to further examine time of a woman's life (puberty, menopause, etc.) and time of exposure to assess correlations between these drastic changes in estrogen and progesterone that naturally occur in the body, and how ingested EEDs could play a role in either exacerbating or preventing unwanted disease.

6.2 Population-level human health assessments in diet and disease

In Chapter 4, the potential for understanding population-level dynamics of dietary behavior was assessed through a two-year WBE study. To date, prior to the COVID-19 pandemic, WBE had largely only focused on illicit and licit substance use across

populations, although earlier on it was established that WBE could be used for surveillance of biological markers of human disease. It was mentioned in several WBE studies that embarking on new applications for WBE would be important to establish the limitations and ability for WBE to develop population-level health assessments that are within the appropriate context of the population served. Given the previous two chapters including the peculiar interactions between phytoestrogen consumption and disease incidence, it seemed logical and appropriate to begin monitoring phytoestrogens at community-level to attempt to deduce patterns of plant-based dietary behavior. Average phytoestrogen intake in a western society is estimated to fall between 1-3 mg per day per person, however, it is noted there are marked cultural differences in consumption patterns that can increase or decrease this range. For example, it had been noted that Asian-Americans can consume 20 mg or more per day, mostly through isoflavone consumption (genistein and daidzein). These differences are notable when referencing the potential health benefits in prevention of disease, such as breast cancer. Additionally, phytoestrogens are known to interact with microbes that live within the human gut microbiome, producing metabolites that have greater health benefits or higher estrogenic potency than their precursor, or parent, compounds. Thus, a multi-omics approach to understanding these interactions at population-level was employed to gain a holistic view of how a population-level assessment of dietary behavior could be practiced within the context of WBE. A small-catchment within a city was chosen for this pilot-scale study in order to best isolate a community and establish trends. Study duration began in August of 2017 and lasted through July of 2019. The compounds chosen to monitor represent two major classes of phytoestrogens, isoflavones genistein and daidzein, and a metabolite of

lignan consumption, enterolactone. Results indicate that average per capita consumption (mg d^{-1} capita) are slightly higher than the estimated average of 1-3 mg reported in literature, with 4.1 ± 2.2 mg d^{-1} per capita in year one and 5.0 ± 2.3 mg d^{-1} per capita for year two in this study. This is thought to be due to significant differences between demographics within this sewer catchment compared to the U.S. as a whole, with the Asian population in particular being nearly five times greater in the study reported herein. This was also reflected when investigating the difference in lignan consumption and isoflavone consumption, with isoflavones being greater; a pattern not typically seen across the greater U.S. population. This could also be due to the greater amount of the Asian population who consume soy-based foods on a more regular basis than other demographics. Finally, distinct and statistically significant seasonal patterns in consumption ($p \leq 0.01$) were measured between Fall and Winter, and Spring and Summer.

Microbial composition was also assessed through a subset of this data set (January through December 2018) to first test the hypothesis that human gut microbiome interactions could be reflected in a community sewage sample, and further, to test if specific bacterial taxa known to interact with consumed phytoestrogens and produce metabolites investigated in this study (enterolactone and equol) could be detected and measured. These hypotheses were proven to be true as genera such as *Clostridium, Blautia, and Bifidobacterium* were all detected in at least one sample, and were able to be semi-quantified by incorporating 16S rRNA bacterial gene quantitative polymerase chain reaction (qPCR) measurements; resulting in a semi-quantitative abundance L^{-1} of wastewater of each identified microbial taxa.

This was the first study to report changes in dietary behavior across a longitudinal monitoring assessment using WBE, while also the first to detect and attempt to quantify human gut microbial interactions at community-level as it pertains to microbially-relevant taxa that transform consumed phytoestrogens to products that have been shown in literature to report greater estrogenic potential and health benefits. Future work to expand on this would be to test this in a greater sewer catchment, further refining the methods employed and making stronger connections between dietary behavior and measured assessments at population-scale. Through a larger-scale assessment, external data sets can be more comparably assessed, particularly in validating consumption patterns and incidence of relevant diseases such as breast cancer.

In Chapter 5, the COVID-19 global pandemic highlighted significant challenges in surveillance of SARS-CoV-2 and downstream healthcare capacity. It was established early on that implemented traditional methodologies for individualized diagnostics were failing to report new cases at a pace that was comparable to viral infectivity within and between communities. Hospital burden and healthcare professional burnout was growing progressively worse as more infected individuals flooded emergency rooms with little to no warning, and lockdown measures, while necessary to prevent viral spread, were implemented with the best sources of information available at the time, leading to loss of productivity, loss of employment, ultimately leading to great economic burden. Public health strategies were exhausted early on; investing a great deal of time and resources to deploy in an untargeted, widescale fashion. This prompted for alternative strategies to surveillance of SARS-CoV-2/COVID-19, in order to get ahead of viral spread, relieve healthcare operations, and implement targeted interventions.

Wastewater-based epidemiology (WBE) was identified as a prime candidate to accomplish these goals, and in a rapid fashion. Due to the existing infrastructure within the City of Tempe, AZ to monitor for opioid use across the city, laboratory operations were able to quickly pivot to be amenable for virus surveillance. The City of Tempe was divided by established sewer catchments, resulting in 11 sites monitored across the city starting from April, 2020 through March, 2021 (n = 1,556 observations), resulting in a high-frequency (high volume of samples collected) and high-resolution (multiple sites) approach to WBE. Samples were processed and analyzed for SARS-CoV-2, targeting the E gene, and reported to city personnel weekly, with subsequent posting to an unrestricted, online dashboard to promote data transparency and an easily-accessible, actionable data set for public health professionals. Due to the rapid turnaround times and inherent nature of wastewater analysis to be inclusive of all individuals inhabiting any one community, it was hypothesized that the weekly wastewater data would precede newly reported clinical cases by zip code, and at city-level, as well as county-level reports of COVID-related hospitalizations and COVID-related deaths. Additionally, it was hypothesized that wastewater monitoring of the virus would enhance the ability to identify infection hotspots within any given community, prompting targeted interventions to quickly mitigate viral spread and further burden experienced downstream.

Results from this study proved these hypotheses to be true as wastewater-derived measurements measured at the city and zip-code level preceded newly reported clinical cases on average by 8.5 $\pm$ 2.1 days during the first wave of the pandemic in June, 2020. When compared to county-level reports, wastewater preceded newly reported cases, COVID-related hospitalizations and COVID-related deaths by 2, 16, and 18 days,

respectively during the same wave of infection. This indicates the true efficacy of the ability for WBE to serve as an early warning system for viral presence and spread within a community, allowing for greater response and resource capacity. A unique attribute of this study population is the switch from wastewater-derived measurements to a lagging indicator as the pandemic grew on into the third wave (December/January, 2020-1). This is thought to be due to increased testing capacity and efficacy specifically within the City of Tempe as it was affiliated with the major university that dominates this city. The increased capacity for testing and downstream reporting made the reports comparable to one another, however, wastewater data can still provide important information for new infection hotspots, as exhibited here in this study as a small community within a larger catchment was identified to have much higher amounts of SARS-CoV-2 viral load (gene copies d^{-1} in wastewater) than the neighboring communities. This identification lead to near-immediate action in response, with subsequent decrease in viral presence either at or below quantifiable levels. These results and corresponding lessons learned are highly impactful as they can serve to inform future investigation for detecting, monitoring, and responding to future emerging infectious disease outbreaks. Future work could continue to assess the changes of SARS-CoV-2 presence over time, while also examining co-occurrence of other known viruses and diseases (influenza virus). Additionally, continuing to test the efficacy of WBE measurements with individualized diagnostics as it relates to testing fatigue and/or apathy, as well as vaccine efficacy, would be especially prudent as the pandemic grows on into its second year.

This critical exploration of novel avenues utilizing WBE suggests that population-level assessments can serve as a timely, informative, and complementary methodology to

be used in tandem with current methods. The discussed limitations should serve as a platform to address in future work, in addition to potentially unanticipated opportunities in order to continue to innovate within this transdisciplinary space.

RESEARCH ETHICS

The field of wastewater-based epidemiology is currently operating in an exciting time as the COVID-19 global pandemic prompted and accelerated the transition from a niche scientific discipline to a more universally realized approach for population health monitoring. The success of tracking SARS-CoV-2 viral spread in community wastewater in order to inform and improve the public health response has now encouraged an even greater interest in monitoring for other aspects of human health to determine comprehensive population-level health assessments. However, while the use of wastewater for public health monitoring can certainly be perceived as beneficial, there is currently a lack of structure in practice, particularly in considering ethical implications. Thus, these matters are briefly addressed here to continue the conversation in pursuit of defining standard of care in this particular arena.

Regardless of where the sample is collected from (i.e., wastewater treatment plant, catchment-level, building-level, etc.) the field of wastewater-based epidemiology has long been recognized for its ability to offer anonymity to the population served, resulting in a composited sample of hundreds to thousands or more contributing individuals. As the lives of the community members are not disrupted in order to collect pertinent population health information, this method of sample collection is also considered passive and/or non-invasive, and, consequently, also viewed as a benefit in order to capture truer aspects of human daily life. However, the level of sample collection resolution brings forth some deeper philosophical questions and concerns in order to prevent unanticipated panic or ramifications that could negatively impact the field of

WBE; how it is perceived by the community, and thus, determining whether it can continue to be used in a responsible fashion as a tool for public health stewardship.

In Chapter 5, the shift from opioid monitoring to SARS-CoV-2 monitoring within the City of Tempe, AZ was discussed. This development in partnership with city personnel would not have been possible without significant community engagement through several workshops and town halls open to the public. This level of transparency allowed the members of the city to trust the partnership, and understand the reasons behind the technology's implementation. Further, for the purpose of opioid monitoring, it was decided to maintain sample collection at a lower resolution as it was important to prevent unwanted attention or isolation of any one particular community, while still allowing the ability to observe use-trends within the city. In the case for SARS-CoV-2 monitoring, it was imperative to track infection dynamics throughout the city at a pace near-comparable to viral spread in order to effectively implement public health interventions and to minimize community burden downstream (i.e., mask mandates). Therefore, for virus monitoring it was considered more widely acceptable to increase the number of catchments, in order to obtain a highly resolved view of COVID-19 infections. Finally, wastewater-informed data on both viral presence and opioid usage throughout the city are publicly available in an online, interactive dashboard, where community members are invited to view and interact with the data. Other strategies such as public webinars and luncheons have been implemented in order to continue the dialogue about wastewater monitoring throughout the city, allowing the community to be well-informed and feel supported by this approach.

Another aspect that is not commonly discussed yet in the field is cultural competence and sensitivity. Due to the passive and composited nature of sample collection, WBE can be seen as a culturally-acceptable approach to obtain human health information as opposed to traditional methods that require individual specimens (i.e., blood, urine, stool, etc.). However, beyond simply collecting a sample, it is imperative to be mindful of how data are interpreted and presented. When samples are collected from within the sewer infrastructure, it may provide more contextually-relevant information that could be a reflection of specific demographics or socioeconomic status. For public health strategies and interventions, this could be seen as a benefit, as the interventions in question could be tailored to better-fit the community; this was discussed in more detail in Chapter 4 when investigating dietary differences between different cultural groups. However, it is important to abstain from taking the information too far out of context or to make overreaching or perhaps, misguided assumptions or to draw inconsiderate conclusions, which could be viewed as improperly targeting specific subpopulations.

As the field continues to broaden into novel areas of public health monitoring, which may prompt more innovative approaches to sample collection and data interpretation, it is important to continue this discourse of considering ethical implications in order to configure a framework for future investigation, as educated by the lessons learned in the work presented herein.

REFERENCES

1. Forouhi, N.G., et al., *Dietary and nutritional approaches for prevention and management of type 2 diabetes.* The BMJ, 2018. **361**.

2. Shim, J.-S., K. Oh, and H.C. Kim, *Dietary assessment methods in epidemiologic studies.* Epidemiology and health, 2014. **36**: p. e2014009-e2014009.

3. Prevention, C.f.D.C.a. *National Center for Chronic Disease Prevention and Health Promotion*. Surveillance Systems 2021; Available from: https://www.cdc.gov/chronicdisease/data/surveillance.htm.

4. Knüppel, S., K. Norman, and H. Boeing, *Is a Single 24-hour Dietary Recall per Person Sufficient to Estimate the Population Distribution of Usual Dietary Intake?* The Journal of Nutrition, 2019. **149**(9): p. 1491-1492.

5. Boogaerts, T., et al., *Current and future perspectives for wastewater-based epidemiology as a monitoring tool for pharmaceutical use.* Science of The Total Environment, 2021. **789**: p. 148047.

6. Choi, P.M., et al., *Wastewater-based epidemiology biomarkers: Past, present and future.* TrAC Trends in Analytical Chemistry, 2018. **105**: p. 453-469.

7. Rousis, N.I., et al., *Wastewater-based epidemiology to assess pan-European pesticide exposure.* Water research, 2017. **121**: p. 270-279.

8. Thomas, K.V. and M.J. Reid, *What else can the analysis of sewage for urinary biomarkers reveal about communities?* Environmental science & technology, 2011. **45 18**: p. 7611-2.

9. Ojo, O., *Nutrition and Chronic Conditions.* Nutrients, 2019. **11**(2).

10. Lichtenstein, A.H., et al., *Perspective: Design and Conduct of Human Nutrition Randomized Controlled Trials.* Advances in Nutrition, 2020. **12**(1): p. 4-20.

11. Weaver, C.M. and J.W. Miller, *Challenges in conducting clinical nutrition research.* Nutr Rev, 2017. **75**(7): p. 491-499.

12. Eastern Research Group, I., *Examination of Clinical Trial Costs and Barriers for Drug Development*. 2014: Department of Health and Human Services.

13. Franco, R.Z., et al., *Popular Nutrition-Related Mobile Apps: A Feature Assessment.* JMIR Mhealth Uhealth, 2016. **4**(3): p. e85.

14. Kristal, A.R. and J.A. Satia, *CHAPTER 9 - Evaluation of Nutrition Interventions*, in *Nutrition in the Prevention and Treatment of Disease*, A.M. Coulston, C.L. Rock, and E.R. Monsen, Editors. 2001, Academic Press: San Diego. p. 123-138.

15. Poslusna, K., et al., *Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice.* British Journal of Nutrition, 2009. **101**(S2): p. S73-S85.

16. Prentice, R.L., et al., *Evaluation and Comparison of Food Records, Recalls, and Frequencies for Energy and Protein Assessment by Using Recovery Biomarkers.* American Journal of Epidemiology, 2011. **174**(5): p. 591-603.

17. Pfeiffer, C.M., et al., *Challenges and Lessons Learned in Generating and Interpreting NHANES Nutritional Biomarker Data.* Adv Nutr, 2017. **8**(2): p. 290-307.

18. Kroll, M., R.K. Phalkey, and F. Kraas, *Challenges to the surveillance of non-communicable diseases – a review of selected approaches.* BMC Public Health, 2015. **15**(1): p. 1243.

19. Thompson, F.E., et al., *The National Cancer Institute's Dietary Assessment Primer: A Resource for Diet Research.* Journal of the Academy of Nutrition and Dietetics, 2015. **115**(12): p. 1986-1995.

20. Thompson, F.E. and A.F. Subar, *Chapter 1 - Dietary Assessment Methodology*, in *Nutrition in the Prevention and Treatment of Disease (Fourth Edition)*, A.M. Coulston, et al., Editors. 2017, Academic Press. p. 5-48.

21. Maurer, J., et al., *The Psychosocial and Behavioral Characteristics Related to Energy Misreporting.* Nutrition Reviews, 2006. **64**(2): p. 53-66.

22. 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.* Journal of the Academy of Nutrition and Dietetics, 2012. **112**(8): p. 1134-1137.

23. Freedman, L.S., et al., *Dealing With Dietary Measurement Error in Nutritional Cohort Studies.* JNCI: Journal of the National Cancer Institute, 2011. **103**(14): p. 1086-1092.

24. Choi, P.M., et al., *Do food and stress biomarkers work for wastewater-based epidemiology? A critical evaluation.* Sci Total Environ, 2020. **736**: p. 139654.

25. Rice, J. and B. Kasprzyk-Hordern, *A new paradigm in public health assessment: Water fingerprinting for protein markers of public health using mass spectrometry.* TrAC Trends in Analytical Chemistry, 2019. **119**: p. 115621.

26. Vitale, D., M. Morales Suárez-Varela, and Y. Picó, *Wastewater-based epidemiology, a tool to bridge biomarkers of exposure, contaminants, and human health.* Current Opinion in Environmental Science & Health, 2021. **20**: p. 100229.

27. 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, 2016. **16**(1): p. 1035.

28. Gracia-Lor, E., et al., *Measuring biomarkers in wastewater as a new source of epidemiological information: Current state and future perspectives.* Environment international, 2017. **99**: p. 131-150.

29. Choi, P.M., et al., *Do food and stress biomarkers work for wastewater-based epidemiology? A critical evaluation.* Science of The Total Environment, 2020. **736**: p. 139654.

30. Choi, P.M., et al., *Social, demographic, and economic correlates of food and chemical consumption measured by wastewater-based epidemiology.* Proceedings of the National Academy of Sciences, 2019. **116**(43): p. 21864-21873.

31. Kasprzyk-Hordern, B., et al., *Estimation of community-wide multi-chemical exposure via water-based chemical mining: Key research gaps drawn from a comprehensive multi-biomarker multi-city dataset.* Environment International, 2021. **147**: p. 106331.

32. 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, 2020. **27**(5).

33. Kitajima, M., et al., *SARS-CoV-2 in wastewater: State of the knowledge and research needs.* Sci Total Environ, 2020. **739**: p. 139076.

34. Bivins, A., Lott, M., Shaffer, M., Wu, Z., North, D., Lipp, E., Bibby, K., *Building-level wastewater monitoring for COVID-19 using tampon swabs and RT-LAMP for rapid SARS-CoV-2 RNA detection.* Preprints, 2021.

35. Betancourt, W.Q., et al., *COVID-19 containment on a college campus via wastewater-based epidemiology, targeted clinical testing and an intervention.* Science of The Total Environment, 2021. **779**: p. 146408.

36. Prado, T., et al., *Wastewater-based epidemiology as a useful tool to track SARS-CoV-2 and support public health policies at municipal level in Brazil.* Water Research, 2021. **191**: p. 116810.

37. Picó, Y. and D. Barceló, *Identification of biomarkers in wastewater-based epidemiology: Main approaches and analytical methods.* TrAC Trends in Analytical Chemistry, 2021. **145**: p. 116465.

38. Prevention, C.f.D.C.a. *Developing a Wastewater Surveillance Sampling Strategy*. 2021; Available from:

https://www.cdc.gov/healthywater/surveillance/wastewater-surveillance/developing-a-wastewater-surveillance-sampling-strategy.html.

39. Venkatesan, A.K., et al., *Assessing the Potential To Monitor Plant-Based Diet Trends in Communities Using a Wastewater-Based Epidemiology Approach*, in *Wastewater-Based Epidemiology: Estimation of Community Consumption of Drugs and Diets*. 2019, American Chemical Society. p. 187-198.

40. Cady, N., et al., *Beyond Metabolism: The Complex Interplay Between Dietary Phytoestrogens, Gut Bacteria, and Cells of Nervous and Immune Systems.* Front Neurol, 2020. **11**: p. 150.

41. Gaya, P., et al., *Phytoestrogen Metabolism by Adult Human Gut Microbiota.* Molecules, 2016. **21**(8).

42. Hughes, R.L., et al., *The Role of the Gut Microbiome in Predicting Response to Diet and the Development of Precision Nutrition Models. Part II: Results.* Advances in Nutrition, 2019. **10**(6): p. 979-998.

43. Tiwari, S.B., et al., *Surveillance of Wastewater for Early Epidemic Prediction (SWEEP): Environmental and health security perspectives in the post COVID-19 Anthropocene.* Environmental Research, 2021. **195**: p. 110831.

44. Bowes, D.A. and R.U. Halden, *Theoretical evaluation of using wastewater-based epidemiology to assess the nutritional status of human populations.* Current Opinion in Environmental Science & Health, 2019. **9**: p. 58-63.

45. Hedrick, V., et al., *Dietary biomarkers: advances, limitations and future directions.* Nutrition Journal, 2012. **11**.

46. Ke, Q., et al., *Association between dietary protein intake and type 2 diabetes varies by dietary pattern.* Diabetology & Metabolic Syndrome, 2018. **10**.

47. Di Maso, M., et al., *Red meat and cancer risk in a network of case-control studies focusing on cooking practices.* Annals of oncology : official journal of the European Society for Medical Oncology, 2013. **24 12**: p. 3107-12.

48. MarÌ-Sanchis, A., et al., *Meat Consumption and Risk of Developing Type 2 Diabetes in the SUN Project: A Highly Educated Middle-Class Population.* PLoS ONE, 2016. **11**.

49. Zhang, C., et al., *Soy product and isoflavone intake and breast cancer risk defined by hormone receptor status.* Cancer Sci, 2010. **101**(2): p. 501-7.

50. Lowcock, E.C., M. Cotterchio, and B.A. Boucher, *Consumption of flaxseed, a rich source of lignans, is associated with reduced breast cancer risk.* Cancer Causes & Control, 2013. **24**: p. 813-816.

51. Iqbal, J., et al., *Dietary isoflavones, the modulator of breast carcinogenesis: Current landscape and future perspectives.* Asian Pacific Journal of Tropical Medicine, 2018. **11**(3): p. 186-193.

52. 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, 2014. **9**.

53. Rietjens, I., J. Louisse, and K. Beekmann, *The potential health effects of dietary phytoestrogens.* Br J Pharmacol, 2017. **174**(11): p. 1263-1280.

54. Ziaei, S. and R. Halaby, *Dietary Isoflavones and Breast Cancer Risk.* Medicines (Basel), 2017. **4**(2).

55. Gracia-Lor, E., et al., *Measuring biomarkers in wastewater as a new source of epidemiological information: Current state and future perspectives.* Environ Int, 2017. **99**: p. 131-150.

56. Gushgari, A.J., et al., *Tracking narcotics consumption at a Southwestern U.S. university campus by wastewater-based epidemiology.* Journal of Hazardous Materials, 2018. **359**: p. 437-444.

57. Daughton, C.G., *Monitoring wastewater for assessing community health: Sewage Chemical-Information Mining (SCIM).* The Science of the total environment, 2018. **619-620**: p. 748 - 764.

58. Rousis, N.I., et al., *Wastewater-based epidemiology to assess pan-European pesticide exposure.* Water Res, 2017. **121**: p. 270-279.

59. Chen, J., A.K. Venkatesan, and R.U. Halden, *Alcohol and nicotine consumption trends in three U.S. communities determined by wastewater-based epidemiology.* Science of The Total Environment, 2019. **656**: p. 174-183.

60. Karimi, Z., *The Association of Meat Consumption and Breast Cancer Risk: A Case Control Study in a Population of Iranian Women.* American Journal of Life Sciences, 2015. **3**(2).

61. Wu, H., et al., *Antioxidant activities of carnosine, anserine, some free amino acids, and their combination.* Journal of Food and Drug Analysis, 2002. **11**(2): p. 148-153.

62. Cross, A.J., et al., *Urinary 1-methylhistidine and 3-methylhistidine, meat intake, and colorectal adenoma risk.* European Journal of Cancer Prevention, 2014. **23**: p. 38U 390.

63. Cross, A.J., J.M. Major, and R. Sinha, *Urinary Biomarkers of Meat Consumption.* Cancer Epidemiology, Biomarkers & Prevention, 2011. **20**: p. 1107 - 1111.

64. Lampe, J., *Biomarkers of Nutritional Exposure and Nutritional Status.* Journal of Nutrition, 2003: p. 956-964.

65. Bhattacharya, A., et al., *The principal urinary metabolite of allyl isothiocyanate, N-acetyl-S-(N-allylthiocarbamoyl)cysteine, inhibits the growth and muscle invasion of bladder cancer.* Carcinogenesis, 2012. **33 2**: p. 394-8.

66. Mie, A., et al., *Human health implications of organic food and organic agriculture: a comprehensive review.* Environmental Health, 2017. **16**.

67. Tsuji, T., et al., *Urinary excretion of vitamin B1, B2, B6, niacin, pantothenic acid, folate, and vitamin C correlates with dietary intakes of free-living elderly, female Japanese.* Nutrition research, 2010. **30 3**: p. 171-8.

68. Woodside, J.V., et al., *Use of biomarkers to assess fruit and vegetable intake.* Proc Nutr Soc, 2017. **76**(3): p. 308-315.

69. Chen, C., et al., *Towards finding a population biomarker for wastewater epidemiology studies.* Science of The Total Environment, 2014. **487**: p. 621-628.

70. Venkatesan, A.K. and R.U. Halden, *Wastewater Treatment Plants as Chemical Observatories to Forecast Ecological and Human Health Risks of Manmade Chemicals.* Scientific Reports, 2014. **4**.

71. Ryu, Y., et al., *Comparative measurement and quantitative risk assessment of alcohol consumption through wastewater-based epidemiology: An international study in 20 cities.* The Science of the total environment, 2016. **565**: p. 977-983.

72. Halden, R.U., *Epistemology of contaminants of emerging concern and literature meta-analysis.* Journal of hazardous materials, 2015. **282**: p. 2-9.

73. Bowes, D.A. and R.U. Halden, *Breast Cancer and Dietary Intake of Endocrine Disruptors: a Review of Recent Literature.* Current Pathobiology Reports, 2019. **7**(3): p. 41-46.

74. Burks, H., et al., *Endocrine disruptors and the tumor microenvironment: A new paradigm in breast cancer biology.* Mol Cell Endocrinol, 2017. **457**: p. 13-19.

75. Diamanti-Kandarakis, E., et al., *Endocrine-disrupting chemicals: an Endocrine Society scientific statement.* Endocr Rev, 2009. **30**(4): p. 293-342.

76. Roy, J.R., S. Chakraborty, and T.R. Chakraborty, *Estrogen-like endocrine disrupting chemicals affecting puberty in humans--a review.* Med Sci Monit, 2009. **15**(6): p. Ra137-45.

77. Reaves, D.K., et al., *Persistent organic pollutants and obesity: are they potential mechanisms for breast cancer promotion?* Endocr Relat Cancer, 2015. **22**(2): p. R69-86.

78. Prevention, C.f.D.C.a. *Basic Information About Breast Cancer*. 2018 [cited 2018 December 11]; Available from: https://www.cdc.gov/cancer/breast/basic_info/index.htm.

79. Geens, T., et al., *A review of dietary and non-dietary exposure to bisphenol-A.* Food Chem Toxicol, 2012. **50**(10): p. 3725-40.

80. Mnif, W., et al., *Effect of endocrine disruptor pesticides: a review.* Int J Environ Res Public Health, 2011. **8**(6): p. 2265-303.

81. Maqbool, F., et al., *Review of endocrine disorders associated with environmental toxicants and possible involved mechanisms.* Life Sci, 2016. **145**: p. 265-73.

82. Agency, E.P. *Bisphenol A Action Plan*. 2010 [cited 2018 December 10]; Available from: https://www.epa.gov/sites/default/files/2015-09/documents/bpa_action_plan.pdf.

83. Giulivo, M., et al., *Human exposure to endocrine disrupting compounds: Their role in reproductive systems, metabolic syndrome and breast cancer. A review.* Environ Res, 2016. **151**: p. 251-264.

84. Authority, E.F.S. *EFSA advises on the safety of paraben usage in food*. 2004 [cited 2018 October 18]; Available from: https://www.efsa.europa.eu/en/news/efsa-advises-safety-paraben-usage-food.

85. Kuhnle, G.G., et al., *Phytoestrogen content of foods of animal origin: dairy products, eggs, meat, fish, and seafood.* J Agric Food Chem, 2008. **56**(21): p. 10099-104.

86. Morgan, M., et al., *Environmental estrogen-like endocrine disrupting chemicals and breast cancer.* Mol Cell Endocrinol, 2017. **457**: p. 89-102.

87. Registry, A.f.T.S.a.D. *Public Health Statement DDT, DDD, and DDE*. 2002 [cited 2018 December 11]; Available from: https://wwwn.cdc.gov/TSP/PHS/PHSLanding.aspx?id=79&tid=20.

88. Patisaul, H.B., *Endocrine disruption by dietary phyto-oestrogens: impact on dimorphic sexual systems and behaviours.* Proc Nutr Soc, 2017. **76**(2): p. 130-144.

89. Alavanja, M.C., M.K. Ross, and M.R. Bonner, *Increased cancer burden among pesticide applicators and others due to pesticide exposure.* CA Cancer J Clin, 2013. **63**(2): p. 120-42.

90. Sungur, Ş., M. Köroğlu, and A. Özkan, *Determinatıon of bisphenol a migrating from canned food and beverages in markets.* Food Chem, 2014. **142**: p. 87-91.

91. Prevention, C.f.D.C.a. *Dichlorodiphenyltrichloroethane (DDT) Factsheet*. 2018 [cited 2018 December 11]; Available from: https://www.epa.gov/ingredients-used-pesticide-products/ddt-brief-history-and-status.

92. Agency, E.P. *Atrazine*. 2018 [cited 2018 December 11]; Available from: https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=209.

93. Medicine, N.L.o. *DDT*. 2018 [cited 2018 December 11]; Available from: https://www.nlm.nih.gov/toxnet/index.html.

94. Patisaul, H.B. and W. Jefferson, *The pros and cons of phytoestrogens.* Front Neuroendocrinol, 2010. **31**(4): p. 400-19.

95. Messina, M., *Soy and Health Update: Evaluation of the Clinical and Epidemiologic Literature.* Nutrients, 2016. **8**(12).

96. Halden, R.U., *Plastics and health risks.* Annu Rev Public Health, 2010. **31**: p. 179-94.

97. Agency, E.P. *DEHP*. 1987 [cited 2018 October 18]; Available from: https://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=14.

98. Chang, Y. and R. Choue, *Plasma pharmacokinetics and urinary excretion of isoflavones after ingestion of soy products with different aglycone/glucoside ratios in South Korean women.* Nutr Res Pract, 2013. **7**(5): p. 393-9.

99. Stanczyk, F.Z. and N.J. Clarke, *Measurement of Estradiol—Challenges Ahead.* The Journal of Clinical Endocrinology & Metabolism, 2014. **99**(1): p. 56-58.

100. Desmawati, D. and D. Sulastri, *Phytoestrogens and Their Health Effect.* Open Access Maced J Med Sci, 2019. **7**(3): p. 495-499.

101. Fletcher, R.J., *Food sources of phyto-oestrogens and their precursors in Europe.* Br J Nutr, 2003. **89 Suppl 1**: p. S39-43.

102. Medawar, E., et al., *The effects of plant-based diets on the body and the brain: a systematic review.* Transl Psychiatry, 2019. **9**(1): p. 226.

103. Moshfegh, A.J., et al., *The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes.* The American Journal of Clinical Nutrition, 2008. **88**(2): p. 324-332.

104. Neuhouser, M.L., et al., *Use of Recovery Biomarkers to Calibrate Nutrient Consumption Self-Reports in the Women's Health Initiative.* American Journal of Epidemiology, 2008. **167**(10): p. 1247-1259.

105. Vuckovic, N., et al., *A Qualitative Study of Participants' Experiences with Dietary Assessment.* Journal of the American Dietetic Association, 2000. **100**(9): p. 1023-1028.

106. Faleye, T.O.C., et al., *Pan-Enterovirus Amplicon-Based High-Throughput Sequencing Detects the Complete Capsid of a EVA71 Genotype C1 Variant via Wastewater-Based Epidemiology in Arizona.* Viruses, 2021. **13**(1).

107. Amoah, I.D., et al., *Effect of selected wastewater characteristics on estimation of SARS-CoV-2 viral load in wastewater.* Environmental Research, 2022. **203**: p. 111877.

108. Faleye, T.O.C., et al., *Wastewater-Based Epidemiology and Long-Read Sequencing to Identify Enterovirus Circulation in Three Municipalities in Maricopa County, Arizona, Southwest United States between June and October 2020.* Viruses, 2021. **13**(9).

109. He, F.-J. and J.-Q. Chen, *Consumption of soybean, soy foods, soy isoflavones and breast cancer incidence: Differences between Chinese women and women in Western countries and possible mechanisms.* Food Science and Human Wellness, 2013. **2**(3): p. 146-161.

110. Messina, M., C. Nagata, and A.H. Wu, *Estimated Asian adult soy protein and isoflavone intakes.* Nutr Cancer, 2006. **55**(1): p. 1-12.

111. Liggins, J., et al., *Daidzein and genistein contents of vegetables.* Br J Nutr, 2000. **84**(5): p. 717-25.

112. Liggins, J., et al., *Daidzein and genistein content of cereals.* Eur J Clin Nutr, 2002. **56**(10): p. 961-6.

113. Clavel, T., et al., *Intestinal bacterial communities that produce active estrogen-like compounds enterodiol and enterolactone in humans.* Appl Environ Microbiol, 2005. **71**(10): p. 6077-85.

114. McCann, S.E., et al., *Enterolignan Production in a Flaxseed Intervention Study in Postmenopausal US Women of African Ancestry and European Ancestry.* Nutrients, 2021. **13**(3).

115. Rodríguez-García, C., et al., *Naturally Lignan-Rich Foods: A Dietary Tool for Health Promotion?* Molecules, 2019. **24**(5).

116. Arai, Y., et al., *Comparison of isoflavones among dietary intake, plasma concentration and urinary excretion for accurate estimation of phytoestrogen intake.* J Epidemiol, 2000. **10**(2): p. 127-35.

117. Franke, A.A., J.F. Lai, and B.M. Halm, *Absorption, distribution, metabolism, and excretion of isoflavonoids after soy intake.* Arch Biochem Biophys, 2014. **559**: p. 24-8.

118. Mayo, B., L. Vázquez, and A.B. Flórez, *Equol: A Bacterial Metabolite from The Daidzein Isoflavone and Its Presumed Beneficial Health Effects.* Nutrients, 2019. **11**(9).

119. Kilkkinen, A., et al., *Use of oral antimicrobials decreases serum enterolactone concentration.* Am J Epidemiol, 2002. **155**(5): p. 472-7.

120. Setchell, K.D. and C. Clerici, *Equol: history, chemistry, and formation.* J Nutr, 2010. **140**(7): p. 1355s-62s.

121. Setchell, K.D. and C. Clerici, *Equol: pharmacokinetics and biological actions.* J Nutr, 2010. **140**(7): p. 1363s-8s.

122. Peterson, J., et al., *Dietary lignans: physiology and potential for cardiovascular disease risk reduction.* Nutr Rev, 2010. **68**(10): p. 571-603.

123. Tomova, A., et al., *The Effects of Vegetarian and Vegan Diets on Gut Microbiota.* Frontiers in Nutrition, 2019. **6**.

124. Driver, E.M., et al., *Alcohol, nicotine, and caffeine consumption on a public U.S. university campus determined by wastewater-based epidemiology.* Science of The Total Environment, 2020. **727**: p. 138492.

125. Iino, C., et al., *Daidzein Intake Is Associated with Equol Producing Status through an Increase in the Intestinal Bacteria Responsible for Equol Production.* Nutrients, 2019. **11**(2).

126. Valentín-Blasini, L., et al., *Urinary phytoestrogen concentrations in the U.S. population (1999–2000).* Journal of Exposure Science & Environmental Epidemiology, 2005. **15**(6): p. 509-523.

127. Zhang, Y., et al., *Urinary Disposition of the Soybean Isoflavones Daidzein, Genistein and Glycitein Differs among Humans with Moderate Fecal Isoflavone Degradation Activity.* The Journal of Nutrition, 1999. **129**(5): p. 957-962.

128. Bowes, D.A., et al., *Unrestricted Online Sharing of High-frequency, High-resolution Data on SARS-CoV-2 in Wastewater to Inform the COVID-19 Public Health Response in Greater Tempe, Arizona.* medRxiv, 2021: p. 2021.07.29.21261338.

129. Hart, O.E. and R.U. Halden, *Modeling wastewater temperature and attenuation of sewage-borne biomarkers globally.* Water Research, 2020. **172**: p. 115473.

130. Caporaso, J.G., et al., *Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.* Proceedings of the National Academy of Sciences, 2011. **108**(Supplement 1): p. 4516-4522.

131. Thompson, L.R., et al., *A communal catalogue reveals Earth's multiscale microbial diversity.* Nature, 2017. **551**(7681): p. 457-463.

132. Bolyen, E., et al., *Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2.* Nature Biotechnology, 2019. **37**(8): p. 852-857.

133. Callahan, B.J., et al., *DADA2: High-resolution sample inference from Illumina amplicon data.* Nature Methods, 2016. **13**(7): p. 581-583.

134. McMurdie, P.J. and S. Holmes, *phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data.* PLoS One, 2013. **8**(4): p. e61217.

135. Davis, N.M., et al., *Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data.* Microbiome, 2018. **6**(1): p. 226.

136. Wickham, H., *ggplot2: elegant graphics for data analysis*. Springer International Publishing. 2016.

137. Ritalahti, K.M., et al., *Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple Dehalococcoides strains.* Appl Environ Microbiol, 2006. **72**(4): p. 2765-74.

138. Pope, L., et al., *New Year's res-illusions: food shopping in the new year competes with healthy intentions.* PLoS One, 2014. **9**(12): p. e110561.

139. Society, N.G. *Seasons*. [cited 2022; Available from: https://www.nationalgeographic.org/encyclopedia/season/.

140. Lampe, J.W., *Isoflavonoid and Lignan Phytoestrogens as Dietary Biomarkers.* The Journal of Nutrition, 2003. **133**(3): p. 956S-964S.

141. Bureau, U.S.C. *United States Census Quick Facts*. [cited 2022; Available from: https://www.census.gov/quickfacts/fact/table/US/PST045221#qf-headnote-a.

142. Kuijsten, A., et al., *Pharmacokinetics of Enterolignans in Healthy Men and Women Consuming a Single Dose of Secoisolariciresinol Diglucoside.* The Journal of Nutrition, 2005. **135**(4): p. 795-801.

143. Matthies, A., et al., *Daidzein and genistein are converted to equol and 5-hydroxy-equol by human intestinal Slackia isoflavoniconvertens in gnotobiotic rats.* J Nutr, 2012. **142**(1): p. 40-6.

144. Johnson, A., L. Roberts, and G. Elkins, *Complementary and Alternative Medicine for Menopause.* Journal of Evidence-Based Integrative Medicine, 2019. **24**: p. 2515690X19829380.

145. Lefkowitz, E.S., et al., *State Patty's Day: College Student Drinking and Local Crime Increased on a Student-constructed Holiday.* J Adolesc Res, 2012. **27**(3): p. 323-350.

146. Bolca, S., et al., *Microbial and Dietary Factors Are Associated with the Equol Producer Phenotype in Healthy Postmenopausal Women.* The Journal of Nutrition, 2007. **137**(10): p. 2242-2246.

147. Ungar, Y., O.F. Osundahunsi, and E. Shimoni, *Thermal stability of genistein and daidzein and its effect on their antioxidant activity.* J Agric Food Chem, 2003. **51**(15): p. 4394-9.

148. Gracia-Lor, E., et al., *Measuring biomarkers in wastewater as a new source of epidemiological information: Current state and future perspectives.* Environment International, 2017. **99**: p. 131-150.

149. Castiglioni, S., et al., *Evaluation of uncertainties associated with the determination of community drug use through the measurement of sewage drug biomarkers.* Environmental Science and Technology, 2013. **47**(3): p. 1452-1460.

150. Banta-Green, C.J., et al., *The spatial epidemiology of cocaine, methamphetamine and 3, 4-methylenedioxymethamphetamine (MDMA) use: a demonstration using a population measure of community drug load derived from municipal wastewater.* Addiction, 2009. **104**(11): p. 1874-1880.

151. Zuccato, E., et al., *Changes in illicit drug consumption patterns in 2009 detected by wastewater analysis.* Drug and alcohol dependence, 2011. **118**(2): p. 464-469.

152. van Nuijs, A.L.N., et al., *Sewage epidemiology — A real-time approach to estimate the consumption of illicit drugs in Brussels, Belgium.* Environment International, 2011. **37**(3): p. 612-621.

153. Choi, P.M., et al., *Wastewater-based epidemiology biomarkers: Past, present and future.* Trends in Analytical Chemistry, 2018. **105**: p. 453-469.

154. Halden, R.U., et al., *Tracking harmful chemicals and pathogens using the Human Health Observatory at ASU.* Online Journal of Public Health Informatics, 2019. **11**(1).

155. Rice, J. and B. Kasprzyk-Hordern, *A new paradigm in public health assessment: Water fingerprinting for protein markers of public health using mass spectrometry.* TrAC, Trends in analytical chemistry (Regular ed.), 2019. **119**: p. 115621.

156. Sanche, S., et al., *High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2.* Emerging infectious diseases, 2020. **26**(7): p. 1470-1477.

157. *Monitoring SARS-CoV-2 Through Wastewater-Based Epidemiology and COVID-19 Clinical Testing Data on a Large US University Campus*. 2021.

158. Sherchan, S.P., et al., *First detection of SARS-CoV-2 RNA in wastewater in North America: A study in Louisiana, USA.* The Science of the total environment, 2020. **743**: p. 140621.

159. Gerrity, D., et al., *Early-pandemic wastewater surveillance of SARS-CoV-2 in Southern Nevada: Methodology, occurrence, and incidence/prevalence considerations.* Water research X, 2021. **10**: p. 100086-100086.

160. Gonzalez, R., et al., *COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology.* Water research (Oxford), 2020. **186**: p. 116296-116296.

161. Peccia, J., et al., *Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics.* Nature biotechnology, 2020. **38**(10): p. 1164-1167.

162. Wu, F., et al., *SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases.* mSystems, 2020. **5**(4): p. e00614-20.

163. Tempe, C.o. *Fighting Opioid Misuse by Monitoring Community Health*. 2021 [cited 2021 15 May]; Available from: https://www.arcgis.com/apps/Cascade/index.html?appid=92073d7f6a6a498b987f2afdab1b9471.

164. Tempe, C.o. *Opioid Dashboard*. 2019; Available from: https://arcg.is/ey0Ha.

165. Staff, S.P. *1 case of coronavirus confirmed in Maricopa County, health officials say*. 2020 [cited 2021 1 June 2021]; Available from: https://www.statepress.com/article/2020/01/spcommunity-coronavirus-in-maricopa-county-asu.

166. Maricopa-County. *Epidemic Curve*. 2021 [cited 2021; Available from: https://phdata.maricopa.gov/Dashboard/e10a16d8-921f-4aac-b921-26d95e638a45?e=false&vo=viewonly.

167. General, U.S.P.S.O.o.I., *The Untold Story of the ZIP Code*. 2013, United States Postal Service. p. 34.

168. Ahmed, W., et al., *SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study.* The Science of the total environment, 2021. **761**: p. 144216-144216.

169. Medema, G., et al., *Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands.* Environmental science & technology letters, 2020. **7**(7): p. 511-516.

170. Corman, V., et al., *Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR.* Euro surveillance : bulletin européen sur les maladies transmissibles, 2020. **25**(3).

171. Ahmed, W., et al., *Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater.* The Science of the total environment, 2020. **739**: p. 139960-139960.

172. Holland, L.A., et al., *An 81-Nucleotide Deletion in SARS-CoV-2 ORF7a Identified from Sentinel Surveillance in Arizona (January to March 2020).* Journal of virology, 2020. **94**(14).

173. Bureau, U.S.C. *QuickFacts -Guadalupe town, Arizona*. 2019 [cited 2021 16 March]; Available from: https://www.census.gov/quickfacts/guadalupetownarizona.

174. Hart, O.E. and R.U. Halden, *Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges.* Science of the Total Environment, 2020. **730**.

175. Nemudryi, A., et al., *Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater.* Cell Reports Medicine, 2020. **1**(6): p. 100098-100098.

176. Wurtzer, S., et al., *Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in waste water, Greater Paris, France, 5 March to 23 April 2020.* Euro surveillance : bulletin européen sur les maladies transmissibles, 2020. **25**(50).

177. Centazzo, N., et al., *Wastewater analysis for nicotine, cocaine, amphetamines, opioids and cannabis in New York City.* Forensic Sciences Research, 2019: p. 167.

178. Wilson, E., et al., *Multiple COVID-19 Clusters on a University Campus - North Carolina, August 2020.* MMWR. Morbidity and mortality weekly report, 2020. **69**(39): p. 1416-1418.

179. Fox, M.D., et al., *Response to a COVID-19 Outbreak on a University Campus - Indiana, August 2020.* MMWR. Morbidity and mortality weekly report, 2021. **70**(4): p. 118-122.

180. New-York-Times. *Tracking the Coronavirus at U.S. Colleges and Universities*. 2021 [cited 2021; Available from: https://www.nytimes.com/interactive/2020/us/covid-college-cases-tracker.html.

181. Randazzo, W., et al., *SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area.* Water research (Oxford), 2020. **181**: p. 115942-115942.

182. Wu, S.L., et al., *Substantial underestimation of SARS-CoV-2 infection in the United States.* Nature communications, 2020. **11**(1): p. 4507-4507.

183. Manabe, Y.C., J.S. Sharfstein, and K. Armstrong, *The Need for More and Better Testing for COVID-19.* JAMA, 2020. **324**(21): p. 2153-2154.

184. Liu, Z.H., et al., *Estimated human excretion rates of natural estrogens calculated from their concentrations in raw municipal wastewater and its application.* Environ Sci Pollut Res Int, 2015. **22**(12): p. 9554-62.

185. Miles, F.L., et al., *Plasma metabolite abundances are associated with urinary enterolactone excretion in healthy participants on controlled diets.* Food Funct, 2017. **8**(9): p. 3209-3218.

186. Ritalahti, K.M., et al., *Quantitative PCR Targeting 16S rRNA and Reductive Dehalogenase Genes Simultaneously Monitors Multiple Dehalococcoides Strains.* Applied and Environmental Microbiology, 2006. **72**(4): p. 2765-2774.

APPENDIX A

SUPPLEMENTAL MATERIAL FOR CHAPTER 4

Table 8. Genistein and daidzein content of selected fruits, nuts, cereals, and vegetables (ug/kg dry weight) adapted from [111, 112].

Food Item	Reported Concentration (ug/kg dry wt.)	
	Daidzein	Genistein
Fruits		
Apricots, dried	50	nd
Clementines	27	270
Cranberries	51	213
Currants	560	2167
Dates, dried	18	54
Figs, raw	28	77
Figs, dried	19	45
Fruit cocktail in syrup	nd	17
Mango, raw	251	212
Mango in syrup	39	54
Melon, Canteloupe	nd	42
Melon, Honeydew	151	117
Passion fruit	245	403
Peaches in syrup	32	59
Pears in syrup	10	52
Plums, raw	5	551
Prunes, dried, raw	52	104
Raisins, California	690	1458
Strawberry, raw	45	457
Strawberry in syrup	nd	223
Nuts		
Brazil nut	12	nd
Chestnuts, raw	79	59
Coconut, fresh	128	185
Hazelnuts	58	194
Peanut Butter	Nd	98
Peanuts, fresh	77	158
Peanuts, dry roasted	37	172
Sesame seeds	37	17
Cereals		
Nestle Shredded Wheat	372	760
Kelloggs Start	101	87
Crispbread, multigrain	6085	5788
Crispbread, rye	25	tr
Crispbread, wheat	66	179

Crispbread, wholemeal wheat	92	130
Biscuits, McVities Chocolate Homewheat	165	60
Biscuits, McVities, Cheddars	5	76
Biscuits, Jacobs Choice Grain Crackers	nd	61
Biscuits, McVities Digestives	259	288
Biscuits, McVities Ginger Nuts	88	41
Biscuits, Jacobs Fig Rolls	140	167
Biscuits, Rakusen's Matzos	4	75
Biscuits, Custard Creams, Golden Biscuit Co.	290	113
Biscuits, McVities Rich Tea Biscuits	132	149
Vegetables		
Potatoes, new, raw	132	304
Potatoes, new, cooked	55	147
Potatoes, old, raw	28	147
Potatoes, old, cooked	5	37
Baked beans	nd	228
Baked beans, heated	51	201
Beansprouts, mung, raw	$39x10^3$	$68x10^3$
Broad beans, raw	74	59
Butter beans, dried, raw	305	847
Butter beans, dried, cooked	185	65
Chickpeas, whole, dried, raw	475	766
Chickpeas, whole, dried, cooked	nd	578
French beans, raw	1198	3372
French beans, cooked	1151	3075
French sliced beans, frozen, cooked	686	1938
French sliced beans, frozen, raw	479	1362
Haricot beans, raw	131	105
Haricot beans, cooked	186	173
Lentils, red, split, dried, raw	139	84
Lentils, red, split, dried, cooked	50	93
Mung beans, dried, raw	50	106
Mung beans, dried, cooked	154	399
Red kidney beans, raw	191	209
Red kidney beans, cooked	311	221

Runner beans, Dunn IHRM, raw	$23x10^3$	$31x10^3$
Runner beans, raw	$13x10^3$	$15x10^3$
Runner beans, cooked	7080	8860
Soyabeans, dried, raw	$583x10^3$	$838x10^3$
Soyabeans, dried, cooked	$411x10^3$	$839x10^3$
Miso	$594x10^3$	$673x10^3$
Textured vegetable protein, raw	$248x10^3$	$438x10^3$
Peas, fresh, raw	nd	232
Peas, dried, raw	41	144
Peas, dried, cooked	32	381
Peas, frozen, raw	nd	268
Peas, frozen, cooked	nd	215
Peas, processed, tinned	nd	70
Split peas, green, raw	130	347
Split peas, green, cooked	nd	128
Asparagus, raw	36	79
Aubergine (eggplant), raw	8	99
Aubergine (eggplant), cooked	50	105
Sprouting broccoli, raw	69	29
Sprouting broccoli, cooked	72	94
Cabbage, green, raw	tr	tr
Cabbage, green, cooked	tr	tr
Cabbage, red, raw	nd	tr
Cabbage, red, cooked	nd	276
Cabbage, savoy, raw	nd	289
Cabbage, white, raw	nd	32
Cabbage, white, cooked	nd	84
Celeriac, raw	nd	442
Celeriac, cooked	nd	420
Chicory, raw	9	nd
Cucumber, with skin	nd	77
Cucumber, flesh only	nd	62
Mushroom, common, raw	12	209
Mushroom, common, cooked	nd	182
Okra, raw	253	112
Pumpkin	262	nd
Radish, raw	nd	45
Salad onions	nd	187
Sweetcorn, on cob, raw	nd	134
Sweetcorn, on cob, cooked	29	45
Sweetcorn, tinned or frozen, cooked	40	66

Sweetcorn, tinned or frozen, raw	74	59
Tomato, raw	nd	480
Turnip, raw	94	72

Abbreviations: nd: not detected in food item; tr: detected at unquantifiable trace concentration

Table 9. Mass spectrometry transitions for measured phytoestrogens.

Analyte	MS/MS Transition*	DW (s)	DP (V)	EP (V)	CE (V)	CXP (V)
Daidzein	252.7 > 132.3, 90.7	50	-100/-100	-10/-10	-54/-54	-9/-1
Genistein	268.6 > 132.7, 62.7	50	-90/-90	-10/-10	-44/-60	-5/-9
Enterolactone	297.0 > 252.9, 106.6	50	-80/-80	-10/-10	-28/-40	-5/-17
Genistein-d4	272.7 > 136.9, 62.8	150	-80	-10	-44	-5
Equol	241.1 > 120.8, 118.7	150	-55/-55	-10/-10	-20/-30	-9/-5

*Parent ion > Quantification ion, Confirmation ion; DP: Declustering potential; EP: Entrance Potential; CE: Collision Energy; CXP: Collision Cell Exit Potential; DW: Dwell Time; V: Volts

Table 10. Method details for investigated phytoestrogen analytes.

Method Duration	*Mobile Phase A*	*Mobile Phase B*	*Flow Rate*	*Injection Volume*
12.07 min	Water, LC-grade	Methanol, LC-grade	0.5 mL/min	10 uL

Table 11. Recoveries, method detection limits (MDLs), and concentrations of phytoestrogens in raw wastewater.

Analyte	Matrix Spike Recovery (Mean ±SD, %)*		MDL (ng/L)	Concentration in Raw Wastewater (this study) (ug/L; avg (min, max))
	0.5 ug/L	*5 ug/L*		
Daidzein	69±1	81±13	21	3.94 (0.40, 12.35)
Genistein	76±5	94±14	38	1.17 (0.32, 2.69)
Enterolactone	75±6	81±18	49	6.49 (0.44, 17.80)
Equol	--	--	120	1.04 (0.13, 2.70)

*Previously reported and validated [39]

Table 12. Correction factors and elimination half-lives (hours) for estimating phytoestrogen consumption.

Analyte	Excretion Factor	Correction Factor	Elimination Half-Life (hrs)	Sources
Daidzein	45%	2.2	3-10	[126, 127]
Genistein	20%	5.0	3-10	[126, 184]
Enterolactone	1.1 mg d^{-1}	1.2	7-17	[142, 185]
Equol	2.7 mg d^{-1}	1.1	7-8	[98, 121, 126]

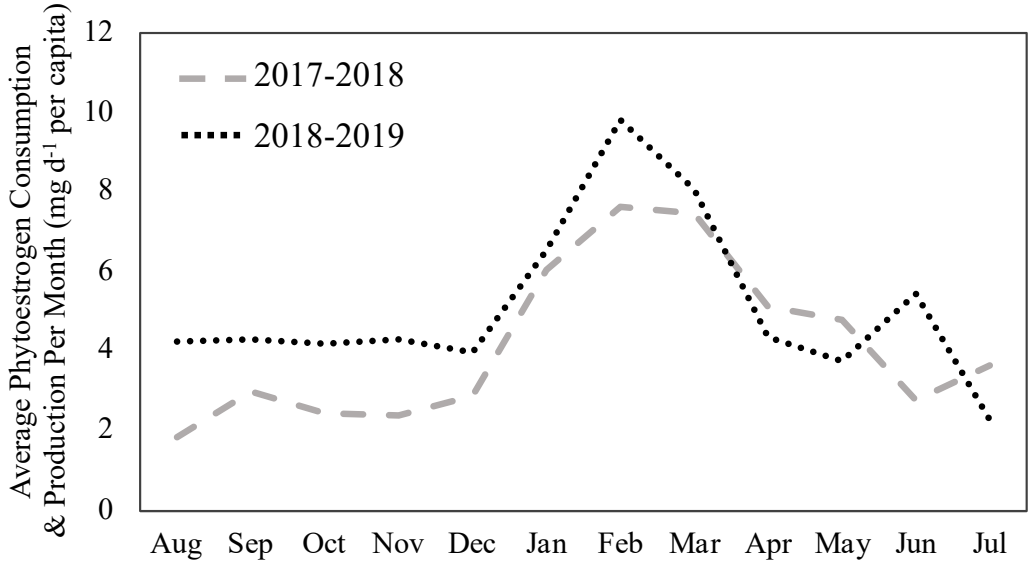

Figure 16. Trends in phytoestrogen consumption and/or production per month from August 2017 through July 2018 (thick grey dashed line) and August 2018 through July 2019 (black dotted line). Each month represents the average of the sum of genistein, daidzein, and enterolactone for each of the seven sample collection days. Results are shown in mg d^{-1} per capita.

Table 13. Monthly wastewater temperatures (°C) shown as minimum, maximum, and average (SD) estimated from historical data in 2008-2010 with recorded ambient temperature in this study from August 2017 through July 2019 as minimum, maximum, and average (SD) for each specific sample collection day per month

Estimated Wastewater Temperature (℃)				Ambient Temperature recorded in this study (℃)				
Month	**Min**	**Max**	**Average (SD)**	**Year**	**Month**	**Min**	**Max**	**Average (SD)**
January	14.1	20.2	17.8 (2.1)	**2017**	**August**	27.5	38.8	33.3 (0.8)
February	14.9	20.9	18.3 (1.9)		**September**	27.8	38.4	33.1 (3.0)
March	17	20.3	19 (1.2)		**October**	19.2	34.1	26.6 (1.7)
April	19.2	25.7	22.2 (1.9)		**November**	15.4	27.9	21.7 (1.7)
May	20	26.9	23.5 (2.2)		**December**	10.5	21.1	15.8 (2.9)
June	22.6	29.3	26.3 (1.3)	**2018**	**January**	7.3	21.5	14.4 (2.3)
July	27.4	31.4	29.1 (1.3)		**February**	5.1	16.1	10.6 (1.5)
August	24.5	30.5	28.9 (2)		**March**	13.8	24.3	19.0 (2.8)
September	26.4	30.4	28.5 (1.5)		**April**	19.0	32.3	25.7 (1.7)
October	20.9	28.3	24.3 (2)		**May**	18.3	30.8	24.6 (4.4)
November	6.9	27.9	15.5 (9.5)		**June**	25.9	40.6	33.2 (0.9)
December	12.1	20.6	17.8 (2.6)		**July**	32.5	44.0	38.2 (1.8)
					August	28.1	39.0	33.6 (1.7)
					September	27.7	41.3	34.5 (0.8)
					October	20.1	27.6	23.9 (2.7)
					November	8.7	22.1	15.4 (1.9)
					December	9.2	18.8	14.0 (1.6)
				2019	**January**	8.2	19.3	13.8 (1.4)
					February	3.3	13.5	8.4 (2.2)
					March	11.6	20.5	16.1 (3.5)
					April	19.9	34.2	27.1 (2.6)
					May	20.0	31.6	25.8 (3.2)
					June	24.3	37.6	30.9 (1.0)
					July	30.0	42.1	36.1 (2.6)

Estimated wastewater temperature previously reported. [129]

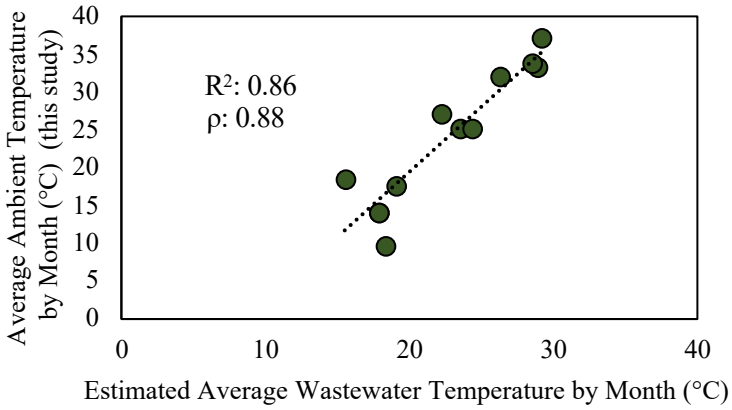

Figure 17. Correlation between wastewater estimates and ambient temperature showing strong linearity and relationship (Spearman's rank-order)

Table 14. Sequences used for 16S rRNA qPCR

Primer and Probe	Sequence (5' – 3')	Source
BAC1055F	ATGGYTGTCGTCAGCT	
BAC1392R	ACGGGCGGTGTGTAC	[186]
BAC1115Probe	FAM-CAACGAGCGCAACCC-TAMRA	

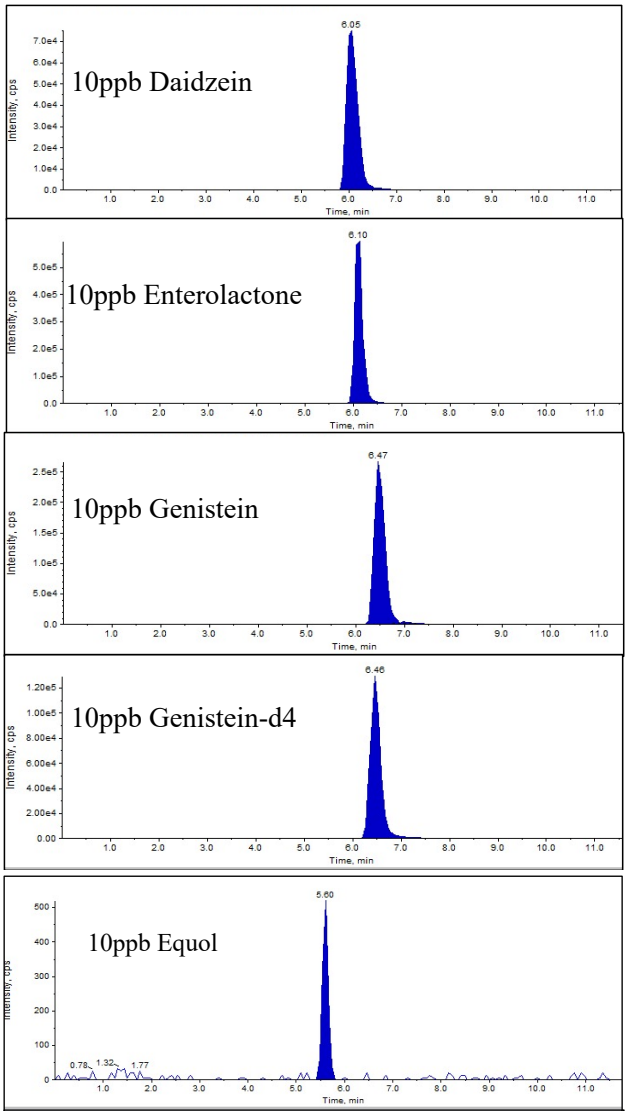

Figure 18. Chromatograms for (top to bottom): Daidzein, Enterolactone, Genistein, Genistein-d4, and Equol; all at concentration of 10 ug/L of standard in water.

Table 15. Calculated p-values using non-parametric Mann Whitney *U* Test with false discovery rate (0.05) corrections using Benjamini-Hochberg Procedure (BH) to test for statistical significance between months, both years of study, weekday versus weekend, and seasonal variability. Statistically significant values noted as (*) for p≤0.05 or (†) for p≤0.01 *Abbreviations: Gen: Genistein; Daid: Daidzein; Ent: Enterolactone; Eq: Equol; Y1: August 2017- July 2018; Y2: August 2018-July 2019; Wk: Weekday (Monday – Friday); Wnd: Weekend (Saturday, Sunday); F: Fall; W: Winter; Sp:Spring; S: Summer.*

Tested Parameter		Gen	BH	Daid	BH	Ent	BH	Sum	BH	Eq	BH
2017	**Aug-Sep**	0.002	0.062	0.167	0.241	0.004	0.017*	0.004	0.033*		
	Sep-Oct	0.035	0.161	0.002	0.014†	0.003	0.017*	0.018	0.052*		
	Oct-Nov	0.064	0.203	0.003	0.014†	0.042	0.073	0.848	0.929		
	Nov-Dec	0.654	0.904	0.568	0.639	0.372	0.464	0.406	0.550		
2018	**Dec-Jan**	0.022	0.146	0.003	0.014†	0.006	0.017*	0.013	0.049*		
	Jan-Feb	0.855	0.992	0.715	0.748	0.004	0.017*	0.167	0.275		
	Feb-Mar	0.715	0.914	0.361	0.489	0.045	0.073	0.465	0.584		
	Mar-Apr	1.000	1.000	0.068	0.142	0.003	0.017*	0.004	0.033*		
	Apr-May	0.668	0.904	0.935	0.935	0.009	0.023*	0.749	0.862		
	May-Jun	0.909	0.992	0.087	0.144	0.087	0.134	0.087	0.155		
	Jun-Jul	0.425	0.652	0.425	0.515	0.017	0.038*	0.017	0.052*		
	Jul-Aug	0.886	0.992	0.003	0.014†	0.032	0.062	0.010	0.047*		
	Aug-Sep	0.100	0.210	0.584	0.639	0.522	0.546	0.273	0.393		
	Sep-Oct	0.088	0.203	0.062	0.142	0.394	0.464	0.935	0.978		
	Oct-Nov	0.085	0.203	0.391	0.500	0.200	0.271	0.180	0.276		
	Nov-Dec	0.949	0.992	0.153	0.235	0.568	0.568	0.482	0.584		
2019	**Dec-Jan**	0.225	0.431	0.018	0.046*	0.142	0.204	0.064	0.134		
	Jan-Feb	0.085	0.203	0.085	0.144	0.006	0.017*	0.025	0.058	0.004	0.013†
	Feb-Mar	0.025	0.146	0.085	0.144	0.025	0.053*	0.025	0.058	0.277	0.277
	Mar-Apr	0.018	0.146	0.004	0.014†	0.003	0.017*	0.002	0.033*	0.007	0.015†
	Apr-May	0.391	0.643	0.004	0.014†	0.423	0.464	1.000	1.000	0.028	0.039*
	May-Jun	0.337	0.596	0.016	0.046*	0.423	0.464	0.078	0.150	0.032	0.039*
	Jun-Jul	0.063	0.203	0.004	0.014†	0.004	0.017*	0.007	0.038*	0.004	0.013†
Year	**Y1-Y2**	0.988	0.988	0.000	0.000†	0.003	0.004†	0.000	0.001†		
Week	**Wk-Wnd**	0.009	0.047*	0.257	0.321	0.080	0.199	0.180	0.300	0.482	0.482
Season	**F-W**	0.000	0.000†	0.000	0.001†	0.000	0.000†	0.000	0.000†		
	W-Sp	0.768	0.945	0.879	0.945	0.114	0.183	0.356	0.517		
	Sp-S	0.000	0.000†	0.000	0.000†	0.012	0.025*	0.000	0.000†		
	S-F	0.975	0.975	0.019	0.034*	0.885	0.945	0.824	0.945		

Table 16. Cost analysis of conducting a WBE study for 24-months at one location

	Item	Cost for 24 mo. WBE Study
Consumables	Sample bottles	$168
	Serological Pipettes	$504
	HLB cartridges	$2,040
	Solvents	$1,224
	2 mL vials	$420
	4 mL vials	$327
	200 µL pipette tips	$1,032
	1000 µL pipette tips	$516
	5 ml pipette tips	$130
	Gloves	$15
	Standards	$36
Consumables Subtotal		**$5,912**
Labor (Lab)	50 h/mo. @ $20/h	$24,000
SUBTOTAL		**$29,912**
Equipment (Lab)	Freezer (-20°C)	$1,000
	Chemical Fume Hood	$7,000
	Autotrace (Dionex; 6-port)	$40,000
	Vacuum Manifold	$1,200
	Reacti-Term Module	$2,250
	HPLC column/guard column	$1,400
	(2) Computer + monitor	$1,200
	Biological Safety Cabinet	$10,000
	LC-MS/MS (Service contract)	$400,000
Startup Equipment Subtotal		**$464,050**
Equipment (Field)	Refrigeratored Sampler	$6,500
	Laser Flow meter	$13,000
	Flow monitoring software	$2,000
Labor (Field)	Field: 10 h/mo. @ $20/h	$4,800
TOTALS WITH STARTUP EQUIPMENT COSTS		
TOTAL WBE STUDY COST		**$520,262**
Estimated Monthly Cost		**$21,677**
Estimated Per Person Cost*		**$2.16**
TOTALS FOR EXISTING LABORATORY SETUP		
TOTAL WBE STUDY COST		**$56,212**
Estimated Monthly Cost		**$2,342**
Estimated Per Person Cost*		**$0.23**

**Based on average population of 10,000*

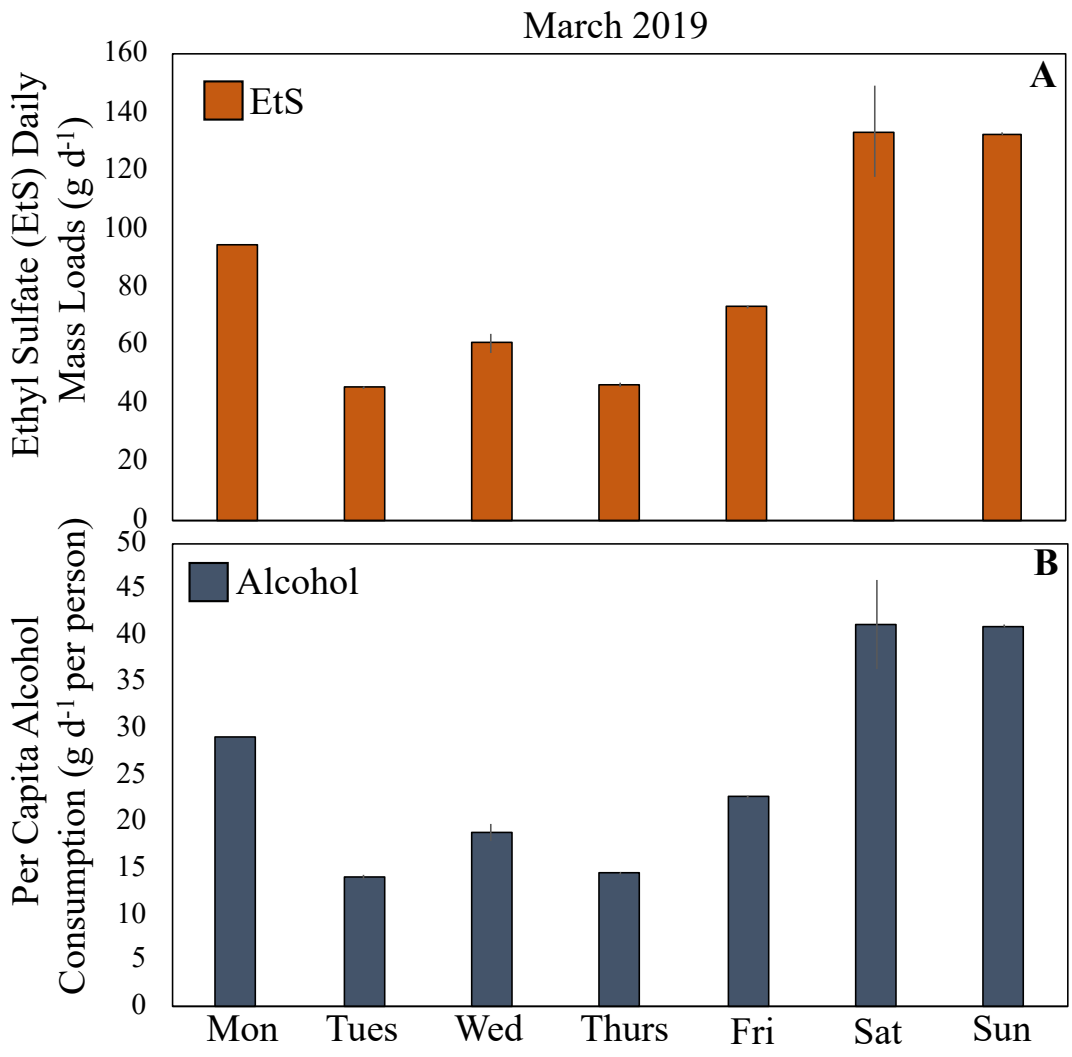

Figure 19 Estimating alcohol consumption over the St. Patrick's Day holiday, showing increases on the 16th (Saturday) and 17th (Sunday). **(A)** Daily mass loading in raw wastewater (g d^{-1}) of ethyl sulfate (EtS). **(B)** Per capita alcohol consumption (g d^{-1} per person). Error bars represent the minimum and maximum results of duplicate samples.

APPENDIX B

SUPPLEMENTAL MATERIAL FOR CHAPTER 5

Table 17. Population estimates for the seven catchments in Tempe and Guadalupe. Tempe estimates were based on 2010 census data, employment data from the Maricopa Association of Governments, and Arizona State University student population. Note that Tempe St. Luke's Hospital was not included here as patient and provider information were not provided.

Year	Month	Area 1	Area 2	Area 3	Area 4	Area 5	Area 6	Area 7	Guadalupe
2020	April	117,769	47,882	42,124	10,498	6,322	7,449	9,382	6,500
	May	127,137	53,788	47,333	13,582	7,736	7,588	10,763	6,500
	June	129,578	55,327	48,690	14,386	8,104	7,624	11,122	6,500
	July	127,662	54,119	47,625	13,755	7,815	7,595	10,840	6,500
	August	133,533	58,837	48,033	13,997	7,926	9,356	12,372	6,500
	September	137,354	61,395	50,076	15,206	8,480	9,423	12,984	6,500
	October	138,664	62,272	50,776	15,621	8,670	9,446	13,193	6,500
	November	139,537	62,857	51,243	15,898	8,797	9,461	13,333	6,500
	December	133,855	63,406	51,068	15,794	8,749	8,686	11,753	6,500
2021	January	138,060	59,551	51,170	15,854	8,777	9,766	11,898	6,500
	February	141,574	61,817	53,067	16,978	9,292	9,830	12,406	6,500
	March	142,520	64,521	53,578	17,280	9,430	9,848	12,542	6,500

Figure 20. Daily flow wastewater measurements for Tempe catchments. Data are only shown for days when samples were collected. Guadalupe and the Tempe St. Luke's hospital location used historical flow data, 370,000 L d^{-1} and 105,992 L d^{-1} respectively and are not shown here.

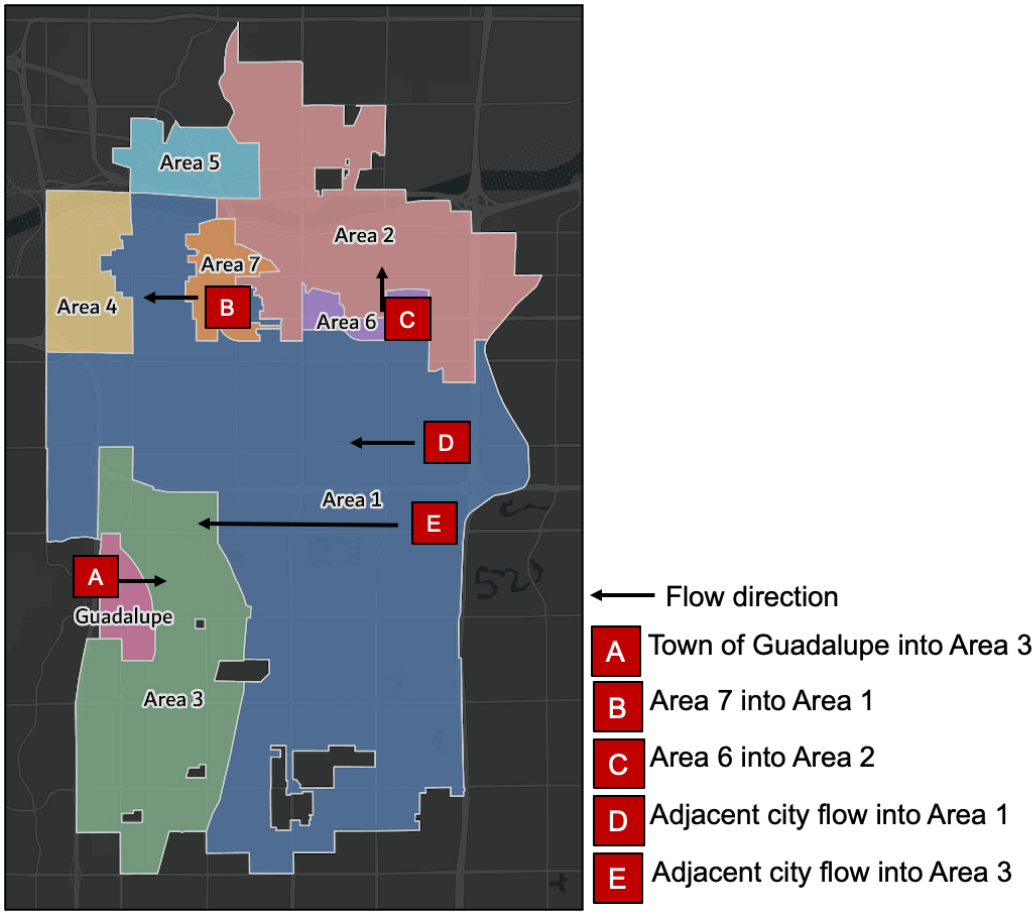

Figure 21. Wastewater flow direction in target catchment areas that illustrate the comingling of waters either from other Tempe-specific catchments or adjacent city flow. Relevant locations include Areas 1, 2, and 3.

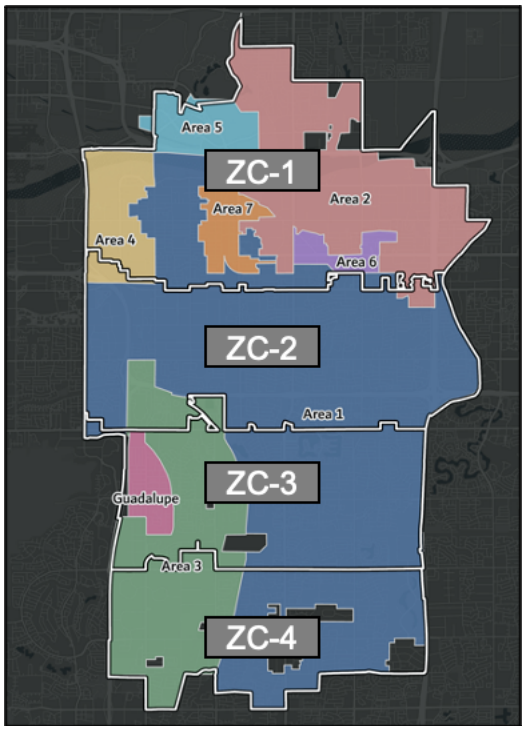

Figure 22. Wastewater catchment areas overlapped with Tempe, AZ zip codes 85281, 85282, 85283, and 85284 correspond to ZC-1, ZC-2, ZC-3, ZC-4 informing percent contribution.

BIOGRAPHICAL SKETCH

Devin Ashley Bowes was born in Bryn Mawr, Philadelphia, Pennsylvania on November 28th, 1990. She attended Upper Merion Area Middle School and High School in King of Prussia, Pennsylvania, and graduated in 2009. She began her undergraduate studies at West Chester University of Pennsylvania in August of 2009, where she was awarded a scholarship to play NCAA Division II Women's Lacrosse. Devin began her undergraduate studies as a Health and Physical Education major with an interest in Adapted Physical Education; working with students with special needs. However, after she completed a nutritional sciences course, she decided to switch into the Nutrition and Dietetics program, and graduated with her Bachelor's of Science (B.S.) degree in August of 2014 with *Summa Cum Laude* honors. Devin later worked in the field as a nutrition counselor at a local Women, Infants, and Children (WIC) clinic in Phoenix, Arizona, where she realized her passion for community nutrition and public health. She decided to pursue a doctoral education, and entered the Biological Design (Ph.D.) program within the Ira A. Fulton Schools of Engineering at Arizona State University in August of 2018. During her tenure at ASU, Devin volunteered as a Co-Instructor for a Fall Seminar Course, as well as served a two-year term as a Biological Design Graduate and Alumni Affairs Committee Member, where she was instrumental in spearheading the program's first Seminar Series. Her dedicated service earned her several awards including a University Graduate Fellowship and Outstanding Graduate Research Award (2021). Devin also served as a graduate student mentor to several undergraduate mentees both in the lab and in their studies. In Fall of 2021, she was awarded the Dean's Dissertation Award by the Ira A. Fulton Schools of Engineering for her innovation and excellence in her dissertation research. Devin is pursuing a career in academia where she aims to conduct independent research and continue to mentor students, while also advocate for diversity and inclusion in academia and research; serving as a role model for women in STEM as well as the LGBTQ+ community.