# Monitoring wastewater for assessing community health: Sewage Chemical-Information Mining (SCIM)

## Christian G. Daughton

*Environmental Futures Analysis Branch, Systems Exposure Division, National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, 944 East Harmon Avenue, Las Vegas 89119, NV, USA*

## HIGHLIGHTS

- Update on Sewage Chemical-Information Mining (SCIM) for assessing public health
- Additional endogenous biomarkers are proposed for targeted monitoring with SCIM.
- Challenges for SCIM include confounding aspects of sewage and data normalization.
- A new concept is proposed that avoids the need for data normalization.
- Biomarkers are needed that measure not just disease but also health.

## GRAPHICAL ABSTRACT

BioSCIM: sewage chemical-information mining targeted at endogenous biomarkers.
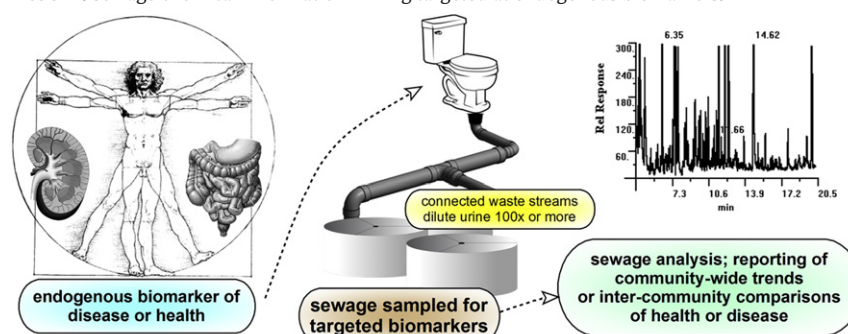
## ABSTRACT

Timely assessment of the aggregate health of small-area human populations is essential for guiding the optimal investment of resources needed for preventing, avoiding, controlling, or mitigating human exposure risks, as well as for maintaining or promoting health. Seeking those interventions yielding the greatest benefit with respect to the allocation of resources is critical for making progress toward community sustainability, reducing health disparities, promoting social justice, and maintaining or improving collective health and well-being. More informative, faster, and less-costly approaches are needed for guiding investigation of cause-effect linkages involving communities and stressors originating from both the built and natural environments. One such emerging approach involves the continuous monitoring of sewage for chemicals that serve as indicators of the collective status of human health (or stress/disease) or any other facet relevant to gauging time-trends in community-wide health. This nascent approach can be referred to as Sewage Chemical-Information Mining (SCIM) and involves the monitoring of sewage for the information that resides in the form of natural and anthropogenic chemicals that enter sewers as a result of the everyday actions, activities, and behaviors of humans. Of particular interest is a specific embodiment of SCIM that would entail the targeted monitoring of a broad suite of endogenous biomarkers of key physiologic processes (as opposed to xenobiotics or their metabolites). This application is termed BioSCIM—an approach roughly analogous to a hypothetical community-wide collective clinical urinalysis, or to a hypothetical *en masse* human biomonitoring program. BioSCIM would be used for gauging the status or time-trends in community-wide health on a continuous basis. This paper presents an update on the progress made with the development of the BioSCIM concept in the period of time since its original publication in 2012, as well as the next steps required for its continued development.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Assessing public health via sewage monitoring—introduction

Robust public health is essential for productive, sustainable communities. The trajectory of public health (the time-trend of a community's health signature) reflects the hazards faced by all individuals, coupled with their collective vulnerabilities to ongoing, daily exposures to myriad types of stressors—exposures spanning the spectrum of socioeconomic, psychologic, physical, and chemical insults.

Although there are many perspectives to how the overall status of collective, community-wide health might be defined, major challenges are faced in how it might be assessed. Measurement and monitoring tools are required to ensure that a community's positive health trajectory can be maintained and that optimal interventions can be taken to mitigate dysfunction, avoid emerging and unrecognized hazards, and reduce the scope of health disparities. As such, the need to quickly detect diminution of collective health requires near real-time monitoring at large scale, all while incurring minimal cost and avoiding the need for human subjects research approvals by institutional review boards (IRBs).

One example of a conventional system designed to surveil and improve public health is the *Healthy People Initiative*. This US federal program was initiated in the late 1970s and undergoes major revision every decade. The Initiative collects health-related data from multiple population scales and sets objectives targeted for improving public health over 10-year time spans. Its current iteration is *Healthy People 2020* (Fielding et al., 2013; ODPHP, 2017).

As a means for assessing public health, the *Healthy People Initiative* relies heavily on self-reporting surveys, which have inherent problems with self-selection and with over- and under-reporting bias. The program also uses comparatively few metrics that rely on chemical-focused biomonitoring of individuals, as biomonitoring is a resource-intensive tool and requires IRB approvals. Moreover, many of the numerous metrics employed by the program are conducive only to infrequent data sampling (e.g., several times per decade).

For the purposes of this article, the *Healthy People Initiative* serves as a backdrop that highlights how a completely new concept for monitoring assessing health at the community scale could contribute a valuable new tool for assessing public health and guiding its improvement. This tool is based on the concept called Sewage Chemical-Information Mining—SCIM (Daughton, 2012a, 2012b). When applied specifically to measuring the overall health status of a community, SCIM becomes a monitoring approach analogous to clinical urinalysis. The difference is that the sample originates from the collective, *en masse* excreta from an entire community. So it essentially views the entire community as an integral "patient". SCIM capitalizes on a biological sample that is continually and readily available in the guise of raw (untreated) sewage. And because of its inherent anonymity, it does not require IRB approvals.

Important to note is that the notion of "community" in the application of SCIM is defined as all individuals using any restroom with a sewage hookup serviced by the same sewage treatment facility. This is the *de facto* population for a sewershed (the sewage catchment area), which necessarily comprises continually varying mixes of residents and visitors alike, as a function of daily population movement among different sewersheds (Daughton, 2012a). Importantly, in this sense, the composition of a "community" is not fixed and may not be repeatable with time. This can become a major limitation in the interpretation of analytical data obtained by any application of sewage-based monitoring. Another ramification from this self-selecting definition of a "community" is that the traditional view of a community may not necessarily align according to sewersheds. Communities that are normally regarded as distinct may belong to the same sewershed, just as a community that might be viewed as well-defined could be located in more than one sewershed.

As a concept that continues to evolve, SCIM holds potential for eventual development into the first continuous-monitoring tool capable of gauging the overall, collective health status of entire communities. The specific application of SCIM discussed in this article is an embodiment called *BioSCIM*, which would rely on the analysis of raw sewage for either of two distinct, general classes of chemical markers. The first class comprises biomarkers of endogenous biochemical processes. These biomarkers would be targeted for their ability to reveal the overall status of health from opposite perspectives: (i) those biomarkers that indicate underlying causes of disease, ill-health, dysfunction, stress, trauma, or injury, and (ii) those that serve as positive (or prognostic) indicators of good health or wellness; note, however, that compared with biomarkers of disease, there are very few biomarkers yet known that can directly gauge positive health (this may be both a cause and a result of why the US healthcare system is oriented toward treatment of disease rather than maintenance and promotion of health). The second general class of targeted chemical markers comprises human metabolites of xenobiotics such as natural products that are associated with either healthy or unhealthy activities or exposures (such as either nutritive constituents or natural contaminants in foods).

A real-time, low-cost public health monitoring system based on BioSCIM applications could foster more attention and commitment at the community level for the need to promote health and reduce exposure to stressors. It could motivate the public, and empower and catalyze communities to take actions tailored to their specific needs for improving collective health or fostering corrective interventions. Indeed, this is an area for which emphasis had been added in the *Healthy People 2020* initiative.

Note that another embodiment of SCIM (often referred to as *sewage epidemiology* or *wastewater epidemiology*) was the first to be developed. It is currently being used to provide more current monitoring data on community-wide use of illicit drugs, drugs of abuse, tobacco, and alcohol (Daughton, 2011). This particular rendition has been gaining increasing use, especially in Europe (Castiglioni, 2016), but it has attracted comparatively much less attention in the US. See Supplementary Table S1 for examples of applications that contrast the differences between sewage epidemiology and BioSCIM.

Note that sewage epidemiology has been applied more recently to indirectly address public health questions. One example was demonstrated by Thomaidis et al. (2016), where community-wide usage rates for various therapeutic pharmaceuticals (psychoactive, antihypertensive, and anti-ulcer drugs), as well as illicit drugs, were acquired via sewage monitoring. Positive correlations were then established with a well-defined period of economic crisis and social strain; some negative correlations were also established, where certain drugs (such as non-steroidal anti-inflammatories and antibiotics) experienced reduced usage (possibly because of reduced ability to purchase health care). And a second example of the application of sewage epidemiology for assessing community response to stress sought correlations in increased daily temperatures with the occurrence in sewage of various drugs and an artificial sweetener used in soft drinks (Phung et al., 2017). Essentially, these xenobiotic chemical markers were used as collective proxies for non-chemical stressors (i.e., socioeconomic hardship and heat stress). By way of contrast, BioSCIM could possibly have been used to target endogenous biomarkers that directly reflected these stresses to see if the same correlative trends could have been more convincingly established.

In general, the monitoring of exogenous xenobiotics such as therapeutic drugs, pesticides, chemicals widely used in consumer products (e.g., phthalates, parabens), tobacco constituents, and alcohol for the purpose of gauging community-wide health faces many limitations and challenges. At best, by targeting human metabolites of these substances, only the incidence of human usage can be assessed—that is, the incidence of exposure. But they can only serve as indirect, imprecise inferences of health or disease because of many complex limitations and confounding variables (see Supplementary Table S2). In contrast, the interpretation of data from endogenous biomarkers is more meaningful and avoids most of the complications that challenge the utility of

monitoring therapeutic drug metabolites. This paper therefore focuses primarily on the prospects of monitoring sewage for endogenous biomarkers that directly reflect disease, stress, or health.

## 2. Objectives

The primary objectives of this article are to:

– Provide an updated overview of the BioSCIM concept for assessing community-wide health.
– Summarize some of the less-recognized aspects of sewage that pose challenges for implementing BioSCIM.
– Present an examination of new candidate biomarkers that merit further evaluation for BioSCIM, especially biomarkers that serve as positive indicators of health (as opposed to disease).
– Present a new conceptual approach for BioSCIM that bypasses the need for data normalization.
– Catalyze interest among researchers from a wide spectrum of disciplines in advancing the development BioSCIM for assessing community-wide health.

Translation of this nascent, transdisciplinary research area into a routine public health monitoring tool could eventually require the expertise from an extremely broad spectrum of disciplines, some of which include:

analytical chemistry, sensor technology, clinical chemistry, clinical medicine and diagnostics, epidemiology (especially molecular epidemiology), public health, human toxicology, biomarker science and pharmacokinetics, wastewater engineering and hydraulics, sewage treatment, sociology/psychology, public policy and risk communication, informatics, statistics, sustainable cities ("smart" cities), environmental justice, urban metabolism (material flow analysis/accounting), modeling, and metabolomics.

Ultimately, a major outcome sought from this research is for SCIM to be recognized as a viable means for examining the status and trends in public health—especially as the first means for the wide-scale, fast, and economical monitoring of the emergence or incidence of disease or stress across entire local communities. SCIM could be viewed as a means of measuring a community's unique, collective health status—akin to a profile, signature, or fingerprint of overall health. SCIM would serve as a form of *en masse* human "biomonitoring". As such, it would overcome some of the limitations of conventional biomonitoring, which by its nature is limited to collecting data only from individuals and incurs very high costs, requires considerable resources, cannot provide results in near real-time, tends to focus on biomarkers of exposure rather than of stress or disease, and requires human subject approvals.

## 3. Update on published advancements relevant to BioSCIM

Compared with research on the use of sewage epidemiology for gauging community-wide use of illicit drugs (e.g., Baker et al., 2014; Castiglioni, 2016; Daughton, 2011; Jones et al., 2014), or more recently, exposure to other anthropogenic xenobiotics such as pesticides (Rousis et al., 2017), comparatively little work directly relevant to BioSCIM has been published in the 5 years since the background, underpinning, and limitations for the original concept was published (Daughton, 2012a, 2012b). To date, few articles have focused on the use of endogenous biomarkers for assessing community-wide health—in particular the archetype class of BioSCIM biomarkers, the isoprostanes (i.e., Chen et al., 2014; Gaw and Glover, 2016; Ryu et al., 2016; Ryu et al., 2015; Santos et al., 2015; Yang et al., 2015). Useful to note is that since 2012, significant additional progress has been published regarding the utility

of isoprostanes as clinical biomarkers of disease; see the recent overviews of van't Erve et al. (2017), Galano et al. (2017), and Milne (2017). BioSCIM as an area of research has also begun to attract some attention from the scientific and lay press, with one story being Arnold (2016).

The aspect of BioSCIM and sewage epidemiology that has attracted the most attention is the need to estimate small-area population size via the use of either chemical markers or biomarkers. New studies on this aspect are limited (i.e., Castiglioni et al., 2013; Chen et al., 2014; Gao et al., 2016; González-Mariño et al., 2017; Lai et al., 2015; Nakada et al., 2017; O'Brien et al., 2017; O'Brien et al., 2014; Rico et al., 2017; Senta et al., 2015; Thai et al., 2014; Zheng et al., 2017); note, however, that non-chemical-based approaches are also being explored (e.g., Thomas et al., 2017). This is important because the overall levels of markers in sewage often must be normalized, and calculation of per capita contributions is one widely accepted approach.

And finally, indications that this field of research may be making progress in establishing its own niche include discussions in various review articles (Eggimann et al., 2017; Fox et al., 2017; Gracia-Lor et al., 2017; Ort et al., 2013).

## 4. Urine as a component of sewage: some perspective

Urine clearly plays a central role in monitoring sewage for human biomarkers. Unfortunately, there are many other contributing flows to domestic sewage that serve to greatly dilute and confound the analysis of this complex matrix. And there are other aspects that offer some unexplored opportunities. Some of these less-discussed aspects are covered in the next sections.

### 4.1. Urine: the chemical messenger of sewage

Urine comprises a vast and largely untapped wealth of chemical information. This is evident in its use in clinical chemistry and medicine, drug development, public health research, forensics, and epidemiology. Chemicals that can be targeted for analysis span the spectrum of endogenous biomarkers reflecting physiological states of stasis, stress, and disease (e.g., proteins, sugars, lipids, and nucleosides), xenobiotics (e.g., pesticides, active pharmaceutical ingredients, and natural product toxins), and countless metabolic transformation products. Biochemicals never suspected of being present in urine are now known to be concentrated in excreted cellular vesicles such as exosomes [see section: Urinary exosomes—potential role in BioSCIM?], presenting yet further opportunities to evaluate the state of human health or chemical exposures from one of the most easily obtained, non-invasive clinical specimens.

Urine in sewage, however, differs dramatically from the urine specimens used in clinical molecular diagnostics. Once urine begins its journey in sewers, it becomes intermixed with other contributors to sewage. Its chemical information content continues to hold value, although continually diminished from the action of physicochemical and biological processes that chemically transform many of the substances comprising the multitudes of natural and anthropogenic chemicals (e.g., Gao et al., 2017). The continual transformation of parent analytes that are targeted by monitoring efforts then complicates the modeling of loadings of analytes that originate from excreta (e.g., McCall et al., 2017).

Urine's dramatic dilution within a much more complex and diluted matrix poses numerous additional analytical challenges compared with clinical analysis. The major additional challenges are the need for much lower limits of analytical detection in the presence of greatly increased background noise and the increased complexities involved with the need to normalize targeted analyte levels to the size of the contributing population so that values can be standardized.

To date, most trace analytical research directed at sewage has explored the potential for targeting metabolites of illicit and licit drugs for the purpose of back-calculating community wide consumption—an

approach that has become known as *sewage epidemiology*. More recently, the scope of this work has expanded to the targeting of various non-drug xenobiotics (where metabolites are preferred targets, so that actual human exposure can be verified), often with the purpose of calculating community-wide per capita exposures (e.g., González-Mariño et al., 2017; Logue et al., 2017; Lopardo et al., 2017).

Important to note is a major distinction between targeting endogenous biomarkers versus markers originating from exogenous chemicals, especially anthropogenic xenobiotics. Endogenous biomarkers serve as direct measures of important physiological processes. They serve as integrative measures of biological effects (effects that can result from exposures to myriad types of stressors), as measures of stressors in their own right, or measures of both. In contrast, the measurement of xenobiotic metabolites reflects exposure only to the parent xenobiotic or its metabolite(s).

Subsequent to its conceptualization in 2001, and beginning in 2005 with its first real-world implementation (by Zuccato et al., 2005), the many facets of sewage epidemiology have been under steady development and refinement (Daughton, 2011). These advancements have primarily been catalyzed by a need to improve upon conventional public survey methodology for gauging drug consumption and for tracking the emergence of new illicit drugs. Sewage epidemiology offers the potential for a faster and less-resource intensive approach, as well as one that can avoid many of the biases and subjectivity associated with conventional public surveys.

With this as brief background, discussed below are several aspects of sewage epidemiology that are infrequently discussed or which continue to pose refractory challenges. One underappreciated aspect is the reality that urine represents only a very small portion of sewage by volume. This greatly amplifies the existing analytical challenges normally faced in urinalysis and introduces some new ones. With respect to the discussion in this paper, note that sewage epidemiology represents one specific embodiment of its umbrella concept—Sewage Chemical-Information Mining (Daughton, 2012a, 2012b). Much of the remainder of this paper will focus on the BioSCIM embodiment, where the target analytes are not xenobiotics, but rather biomarkers of endogenous biochemical processes.

### 4.2. Urine in sewage

At its outset, urine comprises but a very small portion of *yellow water* (depending on the design of the toilet or urinal, and the volume of water used for flushing). But even this initial portion becomes substantially more diluted by overall domestic water use (especially from greywater discharged from sinks, bathing, showers, dishwashers, and clothes washers). Friedler et al. (2013) summarized the daily flows of domestic wastewater (combined blackwater and greywater) across 11 countries. From these data, and using an assumed per capita urinary excretion rate of 1.2 L/day, the percent contribution of urine to the overall flow of domestic wastewater flow was calculated as: a maximum of 1.3% (in Malta), a minimum of 0.54% (in the US), and an average of 0.8% (across 11 countries); in reality, daily per capita urinary output ranges from 0.5 to 1.5 L/day, depending on each individual's hydration status. Urine's contribution to sewage on a volume basis is basically a function of the daily contributions from greywater, which can vary greatly across geographic locales and according to season and socio-cultural norms. A generally accepted figure of roughly 1% is often referenced in the literature (or can be derived from flow data) for the volume of domestic wastewater contributed by urine (e.g., Boutin and Eme, 2016; Landry and Boyer, 2016; Otterpohl, 2002; Simha and Ganesapillai, 2017; Winker et al., 2008).

In reality, however, urine's contribution to sewage volume becomes even more diminished during its transit as domestic sewage. Once domestic wastewaters reach the sewer and merge with other flows, they can be further diluted by infiltration, inflow, industrial inputs, and combined sewer overflow (when present). And ultimately, yet further

stochastic dilution can occur when pumped septage is discharged to the headwaters of treatment plants (a practice increasingly used by many municipalities for disposing of sewage periodically pumped from septic systems).

So once wastewater is sampled at a treatment facility, the urinary contribution of human biomarkers and xenobiotics undergoes extensive and continually varying dilution. This dilution can easily exceed 2 orders of magnitude and thereby increases the need for greater analytical sensitivity and specificity. Analytical methods that are marginal for clinical urinalysis will probably not suffice for sewage without major modifications. The challenge is also greatly increased for design of approaches that ensure representative sampling of bulk sewage flow (Ort et al., 2010).

The entire field of sewage epidemiology clearly relies on the obvious fact that sewage contains urine—however dilute it might be. For this reason, it is worth noting that should technology for urine source-separation eventually become widely employed (e.g., urine-diverting toilets and urinals; see: Lamichhane and Babcock, 2013), the already-low urine content of sewage could be depleted to levels that might fall below the detection capabilities of chemical analysis. On the other hand, if collection of diverted urine were to become centralized (resulting in wastewater with considerably more concentrated urine), then SCIM technologies would remain viable, as the analytical demands could be significantly reduced because the matrix and biomarker levels would more closely resemble those of native, undiluted urine.

### 4.3. Urinary exosomes—potential role in BioSCIM?

In terms of its clinical diagnostic value, blood has historically been favored over urine because of urine's presumed dearth of chemical information. Urine from healthy individuals has been long viewed as comparatively free of macromolecules such as proteins and genetic substances, which hold great potential as molecular biomarkers. This view had been self-fulfilling until the early 2000s because little attention had been devoted to the discovery of new markers in urine. This perspective greatly shifted with the discovery that nanovesicles secreted extracellularly as *exosomes* contain a multitude of markers comprising a wealth of proteins, lipids, glyco derivatives, and genetic material—all originating and greatly enriched from information-generating intracellular processes. These components become concentrated in exosomes, which exhibit surprising stability and persistence during systemic transit and even within urine itself (Mitchell et al., 2009). Urine may actually serve as a far richer source of biomarkers than does blood. The first comprehensive examinations of the urinary proteome began only in the last 10 years.

For the purposes of detecting disruption of biochemical processes (e.g., during stress or development of diseased states), urine might now be viewed as the more useful than blood from which to mine chemical information—if not simply because the body's regulatory mechanisms for maintaining a homeostatic composition of blood is critical for health, and, therefore, changes in biomarker levels in blood can be more difficult to detect (Gao, 2015b). Since urine serves as a major excretory pathway for waste products (which include biomarkers) that are generated by the regulation of homeostasis, it thereby essentially serves to amplify the momentary and otherwise small fluctuations in biomarker concentrations in blood. At least for some biomarkers that have been traditionally monitored in blood, it might be more advantageous to monitor urinary exosomes. This is even more compelling since biomarkers that are otherwise produced at levels below the detection limits of clinical analysis may become sufficiently concentrated within enriched exosomes that their detection becomes possible; note that biomarkers are also significantly expressed on the outer surfaces of exosomes (Hildonen et al., 2016). Background overviews regarding urinary exosomes are widely available (e.g., see: Gao, 2015a; Huebner et al., 2015; Javeed and Mukhopadhyay, 2016; Mora et al., 2016; Nagarajah, 2016; Pant et al., 2012; Street et al., 2017; Vlassov et al., 2012).

With respect to the application of BioSCIM, what roles exosomes might play are currently largely speculative. No research has ever been reported on either the occurrence of intact exosomes in human sewage or on whether ruptured exosomes in sewage might already be contributing meaningfully to the levels of biomarkers in sewage. If undisrupted exosomes persist in sewage, another possibility that could have ramifications for BioSCIM is whether exosomes could be separated and concentrated from sewage, and thereby serve as a physical means for facilitating a highly targeted extraction and concentration of biomarkers from the complex sewage matrix. Methodologies covering a wide spectrum of approaches have been developed for sampling and analysis of exosomes in urine (for one of many possible examples, see: Hildonen et al., 2016). None of these can be considered standard, and none has ever been applied to sewage.

The possible occurrence of exosomes in sewage clearly presents several avenues for potential investigation. Worth noting is that clinical researchers have been working on high-throughput urinary protein identification and quantitation. If a large suite of useful clinical urinary biomarkers could eventually be vetted for BioSCIM, then the potential for rapidly measuring numerous distinct biomarkers in sewage might be achievable via exosome analysis. This clearly could have major implications for the future success of BioSCIM.

## 5. Factors complicating measurement of biomarkers in sewage

The following sections discuss some of the factors that could limit the development of BioSCIM as a viable monitoring tool. An alternative approach is proposed that bypasses these impediments, the major one of which is the need for per capita normalization of data.

### 5.1. Septage—potential contributor to BioSCIM measurement bias

Worth noting is a particular aspect of sewage disposal and handling that could sometimes play a role in biasing SCIM data—particularly for computational models that rely on knowing sewage flow rates and population size. Disposal of pumped septage into WWTPs is a practice that could contribute episodic bias. But the role of septage disposal has not been addressed by any published modeling efforts. Septage is the waste that must be periodically pumped from local septic tanks (including portable toilets) and septic systems; Kookana et al. (2014) present a brief overview of septage handling practices across the world. Historically, septage has been disposed or treated in a number of different ways, but (in the U.S.) it is increasingly being transported to municipal WWTPs, where it is pumped into the headworks.

The amount of septage that might be accommodated by a particular WWTP can vary greatly; the overall additional flow rate is largely a function of the unused design capacity of the WWTP and the composition of the septage (e.g., Bugajski et al., 2016). So the percentage of urine in the sewage flow for a WWTP could vary episodically according to the contributed loading by septage. But moreover, a number of other septage variables could lead to unknown degrees of negative or positive bias with respect to the calculation of biomarker levels (or the levels of xenobiotics such as drugs). The major unknowns could result from: (1) septage diluting the biomarkers in the collective sewage that originates from the community serviced by a WWTP, especially since the age of septage (which can be years) could lead to greatly reduced levels of any biomarkers or xenobiotics that were originally present; (2) septage is sometimes transported outside a geographic locale (so it does not represent the community served by the WWTP); and (3) the volumes of pumped septage can fluctuate greatly with time (a function of weather and season, which dictate when septage can be pumped). The end result from mixing septage with sewage is that: (i) the concentrations of biomarkers or xenobiotics could either be increased or (more likely) reduced, (ii) these intermingled levels originate from separate, discrete populations (only one of which is serviced by the WWTP), and (iii) the levels in the two disparate streams can have origins that are offset in time by years.

### 5.2. Urine dilution and the need for biomarker data normalization

To date, SCIM has been used primarily for determining the per capita usage of drugs across the population serviced by a WWTP. With this type of application, the absolute quantity of the targeted analyte (which, in this application, is usually a drug metabolite) that passes into the WWTP over a sufficient period of time must be quantified; this requires knowing the cumulative flow of the sewage. Also required is a sufficiently accurate estimate of the population served. This permits the normalization of the gross calculated ingested quantity against the population size in order to derive per capita consumption over time. The fact that urine undergoes yet additional, continual, and variable dilution once it enters the sewage stream as diluted yellow flush water poses not just challenges for analysis, but also for data handling—especially with regard to data normalization.

The importance of, and difficulties associated with, quantifying population size in real time are well known (e.g., Bruno et al., 2014; Daughton, 2012a). But little progress has been made with regard to advancements. Efforts to simplify the problem of population normalization have generally relied on the analysis for another chemical marker in sewage that serves as a proxy for population size. The analogous (but simpler) problem exists with the clinical urinalysis, where it has long been recognized that urinary output can vary widely across spot samples and even across 24-hour collections (where volumes can vary by 15-fold; e.g., see: Warrack et al., 2009) as a function of a large number of variables, including hydration, health status, gender, age, body size, fitness, muscle mass, and degree of physical exertion. This in turn can cause wide variation in the absolute concentrations of solutes—both inter-individual and intra-individual. For quantitative data that is statistically valid (facilitating the creation of population reference intervals) and which can be used for population inter-comparisons, this necessitates the normalization of concentrations. Among the most widely accepted denominators for normalizing such clinical urinary data are creatinine levels, osmolality (e.g., via freezing-point depression), and specific gravity. But there are many problems with the use of normalization even in the clinical setting (Daughton, 2012a; Fortin et al., 2008; Ryan et al., 2011; Waikar et al., 2010).

Applying normalization designed for clinical urinary analysis to sewage is fraught with additional problems beyond those faced in the clinic. The use of creatinine in sewage as a proxy measure for population dates back to at least 1976 (e.g., Alexander and Stevens, 1976). While there are many problems surrounding creatinine in particular, analysis of sewage introduces even more complexities—one of the greatest being the unknown but variable half-life of creatinine in sewage as a result of biodegradation (e.g., Brown et al., 1985, 1984), especially degradation by biofilms (O'Brien et al., 2017; Thai et al., 2014).

Important to also note is that when population proxies are based on biomarkers (creatinine being the archetype), an implicit assumption is that the population composition and distribution is similar across communities (for creatinine, especially with regard to age, gender, race, and muscle mass). With this said, the following might be worth considering. Since the variance in intra-individual, daily creatinine output is lower than inter-individual output, creatinine might suffice as a normalization parameter when limited to tracking times trends within a given population—but not for comparisons across populations.

Finally, when applied to urine, osmolality and specific gravity reflect total solute concentration. As such, they also reflect the total endogenous output of metabolism. But as we have seen, little of sewage is composed of urine—a result of receiving inputs from wasted potable water, grey water, sewer infiltration/inflow, treated industrial waste, and sometimes terrestrial runoff and septage. And most of these non-urine flows have dissolved solutes of varying comparisons. So colligative properties for sewage, such as osmolality and specific gravity, would not be useful for comparing across communities.

*5.3. New conceptual approach for BioSCIM: eliminating the need for knowing population size and extent of urine dilution*

A requirement for eventual implementation of BioSCIM is a means for standardizing the levels of the targeted urinary biomarkers measured in sewage. The levels for any given biomarker with an otherwise steady rate of entry to sewage will vary with time according to the marker's initial dilution in urine followed by subsequent dilutions with sewage not originating from urine. To detect changes in trends for biomarker levels for a given community, and especially to compare levels across communities, the levels need to be standardized so that they reflect those that would have otherwise occurred in collective 24-hour urine samples (that is, levels in the absence of the variable dilution that urine undergoes as a result of hydration as well as during its journey in sewage).

There have been two general approaches for normalizing targeted marker levels in sewage for the purposes of SCIM (Daughton, 2012a). The most straightforward is to normalize the biomarker against the estimated size of the *de jure* population deduced from the number of households served by the WWTP. The second is to normalize against a proxy for *de facto* population size; a suitable proxy might be a marker or group of markers that reflect per capita contributions. These are the two approaches that have been used in conventional sewage epidemiology—particularly for calculating average per capita consumption of drugs (e.g., Lai et al., 2015; Rico et al., 2017; Zheng et al., 2017).

Given the extensive research done over the last century on urban metabolism (e.g., material flow analysis—or material flow accounting—one of the basic tools of industrial ecology), it is surprising that no tool yet exists for accurately and quickly measuring the size of the subject populations. With direct measures of population, such as *de jure* census data, a major limitation is that these infrequently obtained data become increasingly out of date, and census data do not necessarily represent *de facto* populations to begin with (Daughton, 2012a). When using a proxy for population size (such as a biomarker), estimates closer to actual *de facto* population are theoretically possible; and the data could be made available in near real-time. But proxy measures are vulnerable to additional limitations, not the least of which is that excreted quantities of most biomarkers (including creatinine) can vary greatly as a function of body mass, age, and gender—and many get entangled by confounding factors (e.g., the origins of creatinine in sewage extend beyond urine) (Daughton, 2012a). But then again, with respect to inter-individual variability, the question can certainly be posed as to whether this type of variability must be accounted for as long as its statistical distribution is similar across populations. Since intra-individual creatinine output is much more constant than inter-individual, creatinine might be sufficient as a normalization parameter to track trends within a given population but not for comparisons across populations.

Bypassing both the need to know the population size and the need to correct for the dilution of urine that varies chaotically within a sewage stream would greatly simplify implementation of BioSCIM for measuring community-wide health. These dual problems could be eliminated with the alternative concept to be described here. Important to note is that this approach would be applicable only to BioSCIM for the purpose of measuring the status of health; it would not provide meaningful results for conventional sewage epidemiology (such as estimating per capita drug usage).

BioSCIM was conceptualized with two major applications in mind: (1) track the relative time trends in collective health within a defined population, and (2) compare the relative status of health across different populations. Since the notion of absolute quantity has no meaning in the context of health, there is no need to know the actual population size. An absolute measure of health or disease does not exist. Only relative comparisons have meaning. Measurements intended to provide a per capita statistic for "health" would therefore not be meaningful.

The following conceptual approach would rely on the use of orthogonal (uncorrelated, independent) biomarkers and their relative levels. It would enable a more elegant solution to the normalization problem, as there would be no need to deal with any of the three major factors that are problematic or time-consuming for conventional sewage epidemiology. First, it bypasses the need to know the sewage flow rate. And with a unitless approach to normalization, it avoids the joint problems of knowing the absolute concentrations of biomarkers (in the originating urine) or the size of the contributing population. These three factors are eliminated with a dimensionless analysis that is inherently self-correcting.

Biomarkers that reflect stress, disease, or injury are most often selected because their levels increase as a function of disease severity (impaired health); their levels in sewage would therefore also increase as a function of the relative size of the contributing sub-population. The concentrations of these biomarkers of disease would be normalized against that of a biomarker whose level trends in the opposite direction—namely, one whose concentration rises with increasing, "positive" health or wellness (prognostic indicators of health). The resultant dimensionless ratio (i.e., "disease:health biomarker ratio"—d:hBR) would eliminate both population size and urine dilution as variables. Importantly, the d:hBR would also serve as a measure that amplifies the collective stress (or health) within a population. Note, however, that a major challenge would still be faced by this concept. A suitable biomarker of "positive" health might prove difficult to find. [see sections: Biomarkers of positive health versus disease and The acute-phase response (APR)]. Biomarkers of health or wellness (*BoH*) can be viewed as those endogenous biochemicals that play roles in the maintenance of homeostasis. This contrasts with biomarkers of stress or disease, which serve to disrupt homeostasis or which are produced by stress or disease.

This approach to "normalization" would essentially rely on the ratio of two orthogonal, endogenous biomarkers—those whose levels move in opposite directions as a function of disease or health. The embodiment described here would use the sewage levels for a biomarker of disease or stress in the numerator (dividend) and the corresponding levels for a positive BoH in the denominator (divisor). [Note: The terminology used here formally refers to fractions rather than ratios. The correct formal names for the two analogous, numeric terms used in a ratio are *antecedent* (instead of numerator), *consequent* (instead of denominator), and *ratio* (instead of quotient)]. The d:hBR would then rise with increasing levels of stress, as it would be accompanied by decreasing wellness. And likewise, a declining ratio would result from reduced stress and increasing wellness. Of course, the antecedent:consequent positions for the two orthogonal biomarkers could be switched as long as consistency is maintained. The resulting ratio when expressed as a dimensionless decimal fraction clearly has no meaning on its own. Its value would reside in tracking its change over time within a given population (establishing trends) and in providing relative, comparative measures between different populations.

Worth pointing out here is an inherent limitation of the d:hBR approach. The resulting data could be confounded if a population comprised two separate and distinct sub-populations—one with overall improving health and the other with overall declining health. In this case, interpretation of the resulting d:hBR values could yield meaningless results or misleading insights.

This new approach for normalization (d:hBR) would not only avoid the problems associated with population size and urine dilution, it would also have a collateral benefit resulting from the fact that by using a ratio of values moving in opposite directions, small changes in disease status would be amplified. An upward or downward change in the numerator would likely be accompanied by an opposing change in the denominator (if the biomarkers are appropriately selected). That is, as the excretion of one marker increases (e.g., the severity or prevalence of a disease increases), the level of the other biomarker (which measures health) could move downward. Note, however, that the

rates of change for the two markers would not necessarily be similar. This is because the markers for disease and for health could have different distributions within a population. For example, a small portion of a population might contribute a large portion of a disease biomarker, while a large portion of the population might contribute the preponderance of the health biomarker. It could also result from the use of biomarkers that show different response sensitivities—where one exhibits a greater absolute change than the other.

While the use of disease:health biomarker ratios resolves some existing major problems, the need for standardization remains. The primary factor requiring standardization is the period of time and portion of flow over which the sewage is sampled—for example, 24-h flow-proportional sampling (Ort et al., 2010). The actual analytical methodologies for the biomarkers would also have to be standardized, and the monitored biomarker levels would have to be expressed in consistent units (e.g., based either on mass or moles of biomarker, even though neither presents an advantage).

Note that one of the confounding problems with interpreting the levels of biomarkers in urine is that disease is not always correlated with higher levels of its markers in urine. Sometimes these levels are negatively correlated with urinary levels. This can occur, for example, when kidney function is impaired (Weaver et al., 2016). This is an important factor in selecting a suitable biomarker.

Worth highlighting is that each ratio using different combinations of biomarker levels would serve as a discrete and independent indicator. A panel of such indicators (using biomarkers that reflect different facets of disease or health) could theoretically be developed to improve the diagnostic and epidemiological value of the population-wide assessments—analogous to the use of conventional, clinical urinary panels.

## 6. Considerations for new potential BioSCIM biomarkers

The following sections present an overview of some of the considerations surrounding the selection of biomarkers for use with BioSCIM. Of particular interest are biomarkers of health.

### 6.1. The acute-phase response (APR)

A specific physiological phenomenon should be noted here because of its relevance to the proposed concept of the disease:health biomarker ratio. The acute-phase response (APR), a core aspect of the innate immune system, is initiated in reaction to physiological stress such as inflammation, infection, or physical trauma. One manifestation of the APR is the dramatic change in levels of the acute-phase proteins (APP). In response to stress, the levels in plasma of one class of APPs can rise (known as a positive acute-phase response) and the levels of another class can decline (known as a negative acute-phase response). The coordinated responses of these two classes are designed to minimize tissue damage and at the same time enhance the processes needed for tissue repair.

In the healthy, non-stressed state, negative acute-phase proteins display elevated normal ranges, and positive acute-phase proteins display depressed normal ranges. In this sense, the negative APPs are examples of "biomarkers of health"—those whose levels are normally elevated in the healthy state.

Examples of well-known positive acute-phase proteins include serum amyloid A (SAA), C-reactive protein (CRP), and haptoglobin (Hp). Some examples of well-known negative acute-phase proteins include albumin, transthyretin (TTR), transferrins (TF), retinol binding protein (RBP), paraoxonase, and cortisol binding globulin. The negative acute-phase response at least partly serves to make available and conserve the amino acids required for elevated and quick synthesis of positive APPs in response to stress.

Unfortunately, few of these negative-APPs are extensively excreted in urine and therefore would not be useful for BioSCIM. And when they are excreted, their urinary levels often follow an inverse relationship to their plasma levels. That is, they tend to be excreted in larger quantities during stress. Instead, the subject of APPs is briefly discussed here for the purpose of highlighting a concept that came to light after the disease:health biomarker ratio (d:hBR) was conceptualized.

In the 1980s, a somewhat analogous concept was developed for veterinary medicine—one that used ratios of the plasma concentrations of various APPs. By dividing the plasma concentration of a positive-APP by that of a negative-APP, the resulting value was recognized as providing a magnified response compared with following the time trend of either marker alone. This ratio was termed the acute-phase index (API)—or acute-phase protein index (APPI). The APPI was intended as a tool that could provide an earlier prognostic indication of disease and thereby better discriminate declining health in animals (or humans)—with higher values indicating greater stress (e.g., Gruys et al., 2005; Martinez-Subiela and Ceron, 2005; Toussaint et al., 2004).

The main difference between the use of APPI as a prognostic indicator versus the disease:health biomarker ratio proposed for BioSCIM is that the APPI uses APP levels in plasma, in contrast to d:hBR which would make use of any classes of markers whose urinary levels move in opposite directions in response to either stress or health. The d:hBR is proposed for BioSCIM also as a means of avoiding the need to know the population size served by a sewage treatment facility or the sewage flow rate. There are two main challenges with respect to locating a candidate biomarker analogous to negative-APPs and which is suitable for BioSCIM. The candidate must be: (1) extensively excreted in urine, and (2) the excreted levels should mirror the plasma levels (i.e., normally elevated levels in a healthy state should be mirrored by elevated urinary levels). A specific example of these challenges is presented in the Supplementary materials for a biomarker called gelsolin.

### 6.2. Are there biomarker alternatives to creatinine for estimating population size or for data normalization?

Given the many caveats surrounding the use of creatinine for normalizing biomarker concentrations (in urine of unknown dilution from hydration), it is worth considering the prospects for an alternative normalization biomarker. The limitations of creatinine are discussed in part [see section: Urine dilution and the need for biomarker data normalization] as well as in Daughton (2012a). The performance parameters for an alternative suitable for use in BioSCIM would have to exceed those of creatinine. As with all biomarkers, some of the major factors contributing to variability of excreted levels and ultimate concentrations in sewage are: biomarker half-life in sewage, health status (e.g., obesity, diabetes), biomarker sensitivity (magnitude of change in quantity excreted as a function of changes in health status), quantity excreted (which dictates requisite method detection limit for analysis), body mass, age, gender (especially the factors of pregnancy and menstruation), diet, ethnicity (e.g., metabolic polymorphisms), and exposure to both therapeutic and illicit drugs.

Most of these variables impact the rate and constancy of biomarker excretion, and ultimately the extent to which excretion varies throughout the day and between individuals. They also conspire to magnify the overall variability in biomarker levels in sewage, thereby degrading the robustness of the marker. This is especially problematic when the intention is to compare biomarker levels across populations.

Not yet known is a biomarker that avoids all of the limitations imposed by these parameters. Indeed, none may exist. Each biomarker exhibits at least one major deficiency with respect to its viability for BioSCIM. With creatinine, for example, the major limitations are daily variations in intra- and inter-individual excretion rates (variations that are higher than generally recognized in the practice of clinical medicine), coupled with reduction in concentration in sewage by unpredictable biodegradation activity (e.g., Thai et al., 2014). This reiterates the importance of selecting a panel of orthogonal biomarkers.

In the course of the search for biomarkers potentially suitable for BioSCIM, two of the new candidates were also identified as potential alternatives to creatinine for normalization: p75 neurotrophin receptor [extracellular domain] and the polyamine $N^1,N^{12}$-diacetylspermine. These are presented in the Supplementary materials (with the caveat that additional major limitations may have been overlooked in the review of the literature—limitations that could nullify the utility of these markers).

### 6.3. Some perspective on the search for endogenous biomarkers suitable for targeting in BioSCIM

Clinical medicine is greatly benefitting from advances in metabolomics, which is being used to identify myriad biochemicals, some of which are possibly suitable biomarkers. These chemicals are usually generated from evolutionarily conserved metabolic pathways. Excreted levels of panels of certain metabolites (as well as the increasing use of their ratios) are used to reveal changes in the biochemical pathway signatures that underlie a variety of pathologies (Hocher and Adamski, 2017). But the vast majority of those biochemicals useful in clinical medicine are not suitable for BioSCIM. With the exception of biomarkers for kidney diseases, many biomarkers are not extensively excreted via urine. But moreover, most also have exogenous, confounding sources that can enter sewage (e.g., many biochemicals occur in an abundance of different foods). And for many (especially larger molecular weight proteins), their stability in sewage is unknown even if they are known to persist in collected urine.

When BioSCIM was formally conceptualized (Daughton, 2012b), a seemingly safe assumption was that a range of useful biomarkers could be gleaned from the extensive published literatures of basic, applied, and translational clinical research (e.g., see: Jain, 2010). But after extensive evaluation of relevant published literature, a realization begins to emerge that despite the countless studies on biomarkers for diagnosis of disease (and for use in prognostic studies), very few studies have used trials of meaningful size, and the findings are often not compelling. This is despite the fact that the U.S. FDA has approved a number of biomarker-based tests; many of these, however, are for biomarkers of very specific diseases (targeted therapy screening) that display very low incidence in the general population. When compared against conventional means of diagnosis, many of these approved markers have yet to yield substantive benefit in clinical medicine. Among the many limitations is the issue of biomarker sensitivity (how much tissue levels or excreted levels change as a function of health or disease severity). At best, the levels of most biomarkers change only a few fold—rarely an order of magnitude—making any signal difficult to distinguish from noise or natural variation. This reality would greatly increase the challenges in establishing time-trends in the sewage for a given community or to compare levels across communities. The field of clinical biomarker research as translated to clinical practice is still in its infancy. The effort required to validate new markers for clinical medicine is a formidable challenge.

As an aside, a clarification is needed here regarding a potential point of confusion. The terms "sensitivity" and "specificity" have completely different meanings in the fields of analytical chemistry and clinical chemistry/diagnostics. Since biomarkers serve as analytical targets in both of these fields, the specific context for the use of both terms is very important. In this paper, the terms sensitivity and specificity are used solely in the context of analytical chemistry, where they refer to an analytical method's ability to detect low concentrations of an analyte and to discern an adequate signal for the analyte in the presence of other types of potentially interfering chemicals. Many of the references cited in this paper, however, involve performance metrics for biomarkers in the setting of clinical chemistry and diagnostics. When examining those references, meanings of these two terms refer to the power of a diagnostic test to correctly identify whether a patient is positive or negative with respect to a disease. Discussions of these important distinctions are available in Saah and Hoover (1997) and in Füzéry et al. (2013); sensitivity and specificity in clinical chemistry are also integral to other concepts used in diagnostics such as the Receiver Operating Characteristic (ROC) Curve and Likelihood Ratios; these two terms and others are frequently used in assessing the discriminatory power and clinical utility of a biomarker.

The progress of research and development for clinical biomarkers is reflected by the following statistics, which provide some perspective regarding the number of endogenous biomarkers that are recognized for their clinical utility versus an estimate of a larger galaxy of potential biomarkers. Using one assessment as an example, with respect to the more than 4,000 secreted proteins known to circulate in the human body, less than 10% (i.e., 375) can be measured reliably. And of these, fewer than half (i.e., 171) have been assimilated into FDA-approved diagnostic tests (Wilson et al., 2016). Accounting for endogenous biomarkers of all chemical classes, the entire human metabolome universe exceeds 100,000 substances (Hocher and Adamski, 2017). The process of validating biomarkers for clinical use (e.g., diagnostics) is very long and challenging (e.g., see overviews: Füzéry et al., 2013; Selleck et al., 2017). If only FDA-approved biomarkers were to be considered, this would clearly pose challenges for locating what would prove to be a much smaller galaxy of biomarkers suitable for BioSCIM. But whether a biomarker has been approved for use by the FDA is not an essential criterion for its use with BioSCIM because the purpose is unrelated to clinical diagnostics. Therefore, the universe of potential markers for BioSCIM is considerably greater than it is for clinical medicine, but the additional restrictions imposed by BioSCIM serve to diminish these numbers. This becomes clear when considering some of the additional requirements imposed by BioSCIM, where a suitable biomarker for BioSCIM must: (i) be extensively excreted via urine, (ii) have molecular stability in sewage, (iii) attain levels that remain above analytical detection limits once urine becomes diluted orders of magnitude in sewage, and (iv) have minimal non-urinary, exogenous origins contributing to sewage (e.g., a biomarker's natural presence in raw foods would confound analysis). The last three factors in particular are not concerns faced in clinical analysis.

There could be numerous biomarkers that might otherwise make suitable candidates to examine for use in BioSCIM. But insufficient information is available to assess them against these criteria. As one of two examples, consider a primary, nonenzymatic glycation product formed from the reaction of DNA with the highly reactive dicarbonyl metabolic byproduct methylglyoxal. This nucleotide advanced glycation endproduct (AGE) of guanosine nucleosides is N2-(1-carboxyethyl)-2′-deoxyguanosine (CedG). Urinary CedG appears to be a high-information-content biomarker, especially for diabetes (Jaramillo et al., 2017; Waris et al., 2015). Despite its receiving increased attention as a useful urinary clinical marker (Gavina et al., 2014; Schneider et al., 2004), it is not clear to what extent it is also formed by (and thereby occurs in) animals and plants that serve as dietary sources. Nor is it clear if it would survive metabolic breakdown in the gut or in sewage, or to what extent it is formed by exogenous sources of methylglyoxal (e.g., whether from ambient exposures or in chemotherapy). And even though CedG was discovered several decades ago, there have been very few population studies on urinary levels; but one study does indicate that levels might vary by 2 orders of magnitude (Schneider et al., 2004). CedG illustrates some of the unknowns and challenges in assessing the utility of a biomarker for BioSCIM.

The second example is presented to illustrate the complexities in assessing a urinary biomarker for its potential utility with BioSCIM. The biochemistry and complex metabolic pathways of "vitamin" D have been examined with increasing rigor for decades; there are myriad reviews (e.g., see: Bartoszewicz et al., 2013; Holick, 2011; Horst et al., 2005). There are two principal natural parent forms of vitamin D—vitamin $D_2$ (ergocalciferol, which occurs in certain foods and serves as the major form in vitamin supplements) and $D_3$ (cholecalciferol, the form produced by UVB irradiation of the skin). Long established is

that these play critical roles in health via certain bioactive metabolites; but not settled is whether low levels of these metabolites result from disease or rather cause disease—or both. The complex, regulated metabolic activation pathways for both forms eventually yield 25-hydroxylated analogs [25(OH)D], which are the specific targets in the clinical analysis of serum for monitoring vitamin D status. But these lipophilic metabolites are not excreted. Through yet more catabolic metabolic steps, two terminal pathways diverge—a lactone pathway and a carboxylic acid pathway. The latter ultimately produces two polar terminal products, calcitroic acid and 1-desoxycalcitroic acid; a portion of this is undoubtedly also excreted as conjugates. It is calcitroic acid (1α-hydroxy-23-carboxy-24,25,26,27-tetranorvitamin $D_3$), the metabolite from calcitriol, that might otherwise serve as a compelling target for BioSCIM. But note that calcitroic acid also results from $D_2$ (Zimmerman et al., 2001). Also note that nearly all studies published on the excretion of calcitroic acid have involved animals; the data from the limited human studies indirectly point to bile as being the more significant excretory route (Avioli et al., 1967; Ledger et al., 1985). There is a bewildering array of unknowns. For example, because clinical medicine has had little interest in excreted metabolites of vitamin D, considerable uncertainty exists with humans as to whether calcitroic acid is primarily excreted via the bile or also via urine—or both (Horst et al., 2005; Reddy and Tserng, 1989; Yu and Arnold, 2016); if it is excreted extensively via the feces, it is unknown how quickly or extensively it would desorb from fecal material into the aqueous phase of sewage. All told, calcitroic acid would not seem to serve as a stoichiometric proxy for biologically active vitamin D status, especially because of continual regulatory shifts in the network of pathways. Other complications are that calcitroate levels can rise or fall with various diseases and that certain vitamin D analog drugs can also contribute to the excretion of calcitroic acid.

With this as background and perspective, the two major unmet needs for development of BioSCIM are: (1) expanding the scope of potentially useful biomarkers that are excreted and which are stable in urine and sewage; the current BioSCIM concept has been limited to the solitary biomarker originally proposed as the archetype—the isoprostane class (Daughton, 2012b), and (2) the need to establish biomarkers that reflect heightened or improved "health" or "wellness" when their levels are elevated (in contrast to biomarkers whose levels are elevated by disease or stress). A related major need is finding biomarkers that can serve as proxies for population size; this is required for the denominator against which biomarker concentrations in sewage are normalized in applications of sewage epidemiology (but not for BioSCIM). Note that regarding the second category (biomarkers of health or wellness), such favorable biomarkers are distinguished from adverse biomarkers of stress in that their levels *increase* with diminishing levels of stress or disease. In other words, instead of wellness being defined by the *absence* of *adverse* biomarkers, it can be defined as the *presence* of *favorable* biomarkers. Developing biomarkers of health (BoH) has proven to be surprisingly difficult in clinical medicine. And the few that have been developed are limited to monitoring serum (as they are not excreted in urine). Having at least one biomarker of health is very important for the implementation of BioSCIM because it can obviate the requirement for knowing the population size, which is a variable extremely difficult to accurately calculate.

## 6.4. Biomarkers of positive health versus disease

"Health"—as opposed to disease—has long proved difficult to define. This is undoubtedly because health is a dynamic process rather than a definitive physiological state. Health involves the capability to continually resist adverse physiological perturbation. It involves "metabolic flexibility" to maintain homeostasis despite continual challenge by an ever-changing spectrum of stressors. This is the realm of the concept of "allostasis". A vast literature has tackled the need for an operational

definition of health. For some examples and more perspective, see Ghini et al. (2015), van der Greef et al. (2013), and Vogt et al. (2016).

Indeed, Vogt et al. (2016) note the following: "As philosopher Hans-Georg Gadamer puts it, 'health itself' is the ability to 'forget that one is healthy'." It should then come as no surprise that one of the paradoxes in the practice of medicine is the difficulty in assessing positive health (wellness or resilience)—as opposed to disease and vulnerability. Assessment of health is almost always inferred indirectly as a default condition involving the likelihood of the *absence* of disease—from the *absence* or reduced levels of markers for disease, stress, dysfunction, or injury. In contrast, the numbers of markers whose levels increase with improving health are extremely limited. And these markers are usually not endogenous biochemicals. Common examples are eustress such as physical conditioning (e.g., exercise performance and duration), consumption of healthy foods, good sleep hygiene, and finding purpose in life. Clinicians have far more tools for directly measuring disease or dysfunction than for health or wellness.

This might partly explain why the practice of medicine is predominantly reactive—with its primary focus on treating disease rather than on its prevention. This state of medical care is beginning to slowly change with the advent of new approaches to the practice of health care that place more emphasis on disease prevention and health promotion—examples being those that incorporate a systems biology approach, such as P4H (e.g., see: Sagner et al., 2016). P4H would represent a revolutionary approach to the practice of medicine with an orientation directed toward health (rather than disease)—an approach that is Predictive, Preventive, Personalized, and Participatory.

Another aspect of potential biomarkers of health is that many may exhibit inverted-U dose-response curves, where initially elevated levels reflect good health, but continued elevation (e.g., where the marker becomes over-expressed) can cause or reflect stress. Levels of biomarkers of health must be maintained within a "normal" range. This contrasts with biomarkers of diseases, which usually follow a rising response curve (eventually plateauing) with increased disease severity. And sometimes, the higher the levels, the worse the prognosis. A prime example of a potential biomarker of health with a biphasic dose response is insulin-like growth factor (e.g., IGF-1) or its associated binding protein (e.g., IGFBP-3). The IGF-axis plays a critical role in regulating cellular growth and survival—in organs throughout the body. Its over-expression, however, is associated with cancers.

Biomarkers can show seemingly incongruous behaviors—sometimes rising and other times falling in response to different stressors or in response to the same stressor among different individuals. The following is but one example of this incongruous, mixed monotonic dose-response behavior. This type of behavior would clearly confound the interpretation of BioSCIM data.

HPMA (3-hydroxypropyl mercapturic acid) is the ultimate metabolite that exits the major detoxification pathway for highly cytotoxic acrolein and which involves the precursor non-enzymatic acrolein–glutathione conjugate. HPMA would ordinarily be expected to track the endogenous, systemic generation of acrolein (along with additional, exogenous contributions from ambient and occupational exposures to acrolein). HPMA is extensively excreted in the urine. HPMA levels become significantly elevated by a number of factors, one of the more prevalent being smoking (Eckert et al., 2011). It also becomes elevated by many other conditions that develop from inflammation. With this said, since HPMA is an end-product of a detoxication pathway for a stressor (acrolein), its levels do not necessarily correlate positively with levels of the stressor that generates acrolein. Even though HPMA results indirectly from a stressor (as with other mercapturic acids), it directly represents a functioning (healthy) detoxification system. So it can be viewed as a marker of both health and disease—that is, a healthy response to a stressor.

If the acrolein-detoxication pathway becomes impaired, HPMA production falls. This can be seen in the case of acute tissue damage, such as stroke, or in diseases involving cognitive impairment (Yoshida et al.,

**Table 1**
Biomarkers meriting further evaluation for use with BioSCIM.[a]

| Biomarker[b] | Example | Physiological role or origin | Limitations/advantages |
|---|---|---|---|
| Isoprostanes (IsoPs) | The IsoPs represent the archetype biomarker for the original conceptualization of BioSCIM. They are not discussed in this paper because they have already been covered in depth (Daughton, 2012b). | Broad class of prostaglandin-like free-radical catalyzed oxidation products from certain polyunsaturated fatty acids (in contrast to the cyclooxygenase-formation of the analogous prostaglandins). A well-known example is 15-$F_{2t}$-IsoP (8-isoPGF$_{2\alpha}$). | IsoPs serve as time-integrative biomarkers of systemic oxidative stress. They are extensively excreted in urine and display excellent chemical stability. Extensive array of analytical methodology as well as clinical studies already exist. But only preliminary studies relevant to BioSCIM have been published. IsoPs merit more study. |
| Desmosines | Comprise two non-natural amino acids, each substituted with four lysine residues. Desmosine (DES: 1,2,3,5-tetrasubstituted pyridinium amino acid) and isodesmosine (IDES: 1,3,4,5-tetrasubstituted pyridinium amino acid). | Serve as cross-links conferring elasticity to the protein elastin. Occur only in elastin, which is essential for flexibility of connective tissues. Only a finite supply in body, as synthesis is repressed after growth ceases. Extensively excreted in urine only upon damage by disease or injury to elastin, as well as from normal aging. | Excretion from normal aging process may obscure accelerated release as a result of disease. Excretion trends downward as elastin supply is depleted in advanced disease states. Unusual biomarker in that the body can only excrete a fixed amount over a lifetime. Very chemically stable. Excretion may be independent of drug therapy. Significant, extraneous dietary sources not known. Considerable published research. |
| Bone Turnover Markers (BTMs) | The principal urinary BTMs are NTX and CTX ["N-terminal telopeptide crosslinks" (amino-terminal collagen crosslinks) and "C-terminal telopeptide crosslinks" (carboxy-terminal collagen crosslinks), respectively] along with the cross-linking pyridinolines [pyridinoline (PYD) and deoxypyridinoline (DPD, or D-PYR)]. | Both of these classes are integral to the physical, cross-linked structure and stability of type I collagen, which is the major constituent of the organic matrix of bone. They are released during bone resorption (breakdown), which is a natural process required for maintaining bone health and strength but whose balance is disrupted by a number of diseases. This leads to elevated urinary BTM levels. | Substantial intra-individual variation in diurnal excretion. It is therefore unclear if collective population-wide BioSCIM data would reveal meaningful trends. Resorption can be exacerbated by a number of drug therapies. Excreted levels may not be influenced by diet. But unknown is whether BTMs in raw or digested foods might be released to sewage and therefore confound BioSCIM data. |
| Pterins | Neopterin is a member of this class of biochemicals that share a 4-keto-2-amino pteridine ring. | Neopterin is a catabolic by-product of cyclic guanosine triphosphate and is generated primarily as a result of activation of the cell-mediated immune system. Produced in large quantities as a result of a wide spectrum of infectious diseases and some diseases caused by—or associated with—excessive inflammation. | Occurs as a natural product in some foods, such as tomatoes, spinach, and beets, which, if entering sewers, could confound excreted levels resulting from endogenous production. Large intra-individual natural daily variance in excretion. Susceptible to photolysis. Chemically stable in urine. Excreted in substantial quantities. Serves to integrate the collective stress from a large number of adverse conditions. |
| mtDAMPs | "Mitochondrial-derived damage-associated molecular patterns" are members of larger classes called "alarmin" or "danger signal" molecules. One in particular is fMLP, the tri-peptide $N$-formylmethionine-leucylphenylalanine. | fMLP is a chemotactic peptide released from mitochondria in response to localized necrosis or serious systemic inflammation. It attracts neutrophils (via surface receptors) and stimulates oxidative bursts. fMLP is an integral part of the innate immune system. It serves as a mimic for the same peptide sequence released by pathogenic bacteria. | fMLP is known to occur in agricultural and house dusts. But it is not yet known if it also occurs in sewage other than via urine (e.g., from dusts or bacterial lysis). An exogenous source could confound its utility as a BioSCIM marker. |
| Polyamines | The aliphatic polyamine with most potential as a biomarker of disease is $N^1,N^{12}$-diacetylspermine (DAS or DiAcSpm). | The aliphatic polyamines (PAs), such as putrescine, spermidine, and spermine are essential for regulatory and control functions for all mammalian cells. DAS is a PA catabolite that is normally well regulated and extensively excreted in urine. It becomes dramatically upregulated with the onset and progression of a number of different cancers. Significantly, DAS exhibits little intra-individual or inter-individual daily variation among healthy individuals. | Diacetyl-PAs are probably exclusively endogenous biomarkers (of cancer, kidney disease, and diabetes, among others). They originate in sewer predominantly from urine, as they are not known to have meaningful dietary sources. PAs are very chemically stable with respect to heat and pH, indicating that they may persist in sewage. DAS should be examined as an alternative to creatinine for the normalization of biomarker levels.[c] |
| Nerve growth factor receptor | The extracellular domain of the transmembrane p75 neurotrophin receptor is called p75$^{ECD}$. | p75$^{ECD}$ is enzymatically cleaved from the p75 neurotrophin receptor during neuronal development, injury, and certain neurological diseases (especially amyotrophic lateral sclerosis). It is extensively excreted in urine but only at very low, constant rates in healthy individuals | p75$^{ECD}$ is systemically shed not just during neuronal injury or disease, but also during pregnancy and the first month of life. It is therefore not just a marker for disease. Furthermore, the incidence of neurological disease in a population might be too low for detection of signal over background. This, however, could make p75$^{ECD}$ an alternative for creatinine in estimation of population size or for biomarker data normalization.[c] |
| Vitamin D-binding protein (VDBP) | A 58-kDa glycoprotein (also known as Gc-globulin) binds with vitamin D metabolites, serving to transport, protect, and recycle them in circulation. Also serves as a scavenger (along with gelsolin) of toxic levels of free actin. | Present at very high levels in the serum of healthy individuals (around 0.4 mg/mL) but correspondingly very low urinary levels. Urinary levels become elevated by several orders of magnitude during development and manifestation of a variety of kidney diseases. | The relative urinary levels between healthy and diseased states is much larger than for most biomarkers. This would greatly enhance the odds of successful VDBP detection in sewage. |
| Monocyte chemoattractant protein-1 (MCP-1) | A 13-kDa cytokine, also known as CC-chemokine ligand 2 (CCL2) or small inducible cytokine A2. Renowned as the most potent chemotactic factor for monocytes. | Becomes over-expressed with renal disease (e.g., diabetic nephropathy). Even though urinary levels of MCP-1 for healthy and diseased states do not differ as much compared with VDBP, they still | Increased urinary levels of MCP-1 can be confounded by acute kidney damage, such as that caused by normal physiological responses to healthy activities, such as strenuous, sustained |

*(continued on next page)*

**Table 1** (continued)

| Biomarker[b] | Example | Physiological role or origin | Limitations/advantages |
|---|---|---|---|
| Gelsolin | Together with VDBP, gelsolin is a ubiquitous, systemic protein that is one of the more important scavengers of actin. | change by over an order of magnitude. Plasma levels in healthy individuals range from 200 to 300 mg/L. Key role in regulating disassembly of actin filament released from cellular injury and which can reach toxic levels. Can therefore become depleted after cellular injury, acting like a negative acute-phase protein. Gelsolin itself is cleaved by caspase-3, releasing the terminal (or truncated) fragment, t-gelsolin (or tGelsolin). | exercise. Most data on urinary levels involve fragments such as t-gelsolin. Little data reported on urinary levels of parent gelsolin. High plasma levels of gelsolin can therefore become translated into low gelsolin fragment levels in the urine, and vice-versa. As with other negative APPs, gelsolin can also be up-regulated in some disease states, so its plasma levels may not be easily interpreted as reflecting health or stress. |

[a] See Supplementary materials for narrative discussions (and supporting references) regarding each biomarker (excluding the isoprostanes).

[b] Individual biomarker or class of related biomarkers.

[c] For further background, see section: Are there biomarker alternatives to creatinine for estimating population size or for data normalization?

2015; Yoshida et al., 2012). Presumably, the production or availability of glutathione is reduced by such damage, resulting in lower levels of acrolein–glutathione precursor for HPMA and thereby the diversion of acrolein away from detoxication and instead toward reactions that lead to physiological damage. So HPMA excreted levels can rise or fall depending on the type or magnitude of stress. When employed in the clinical setting, a given biomarker that can respond differently depending on the type or extent of insult can still be useful when used as an indicator for a uniform patient population with a specific disease. But when subpopulations are naturally mixed with respect to their health status (such as with the collective urine from a local community), the results from BioSCIM may be difficult to interpret for certain biomarkers. Collective excreted levels could involve falling and rising contributions from individuals independent of their health status.

The few biomarkers that could be considered indicators of health (or required for maintenance of health) and which are used in clinical medicine are either not excreted in the urine or they have additional, exogenous sources (e.g., dietary) that would confound the interpretation of their levels in sewage. Well-known examples include serum testosterone (in men), DHEA-S (dehydroepiandrosterone-sulfate), HDL (high-density lipoproteins), and IGF-1 (insulin-like growth factor I); the utility of some of these markers (especially testosterone and DHEA-S) also suffer from the fact that their levels tend to decline with age. The negative-APPs serve as another possible example of "positive" biomarkers of health because their levels should normally be elevated—declining only during stress (e.g., infection, inflammation, or physical trauma). Also see *gelsolin* in Table 1 and in Supplementary materials.

An alternative approach to endogenous biomarkers for gauging health (wellness) would target certain exogenous chemicals, such as those associated with what is generally accepted as "healthy" nutrition. To be useful with BioSCIM, these would generally comprise unique metabolites resulting from the consumption of certain nutritious (functional) foods that are correlated with reduced incidence of disease [see section: Indirect markers of health/wellness: dietary markers].

With respect to BioSCIM, there are few possible candidates to serve as positive indicators of health. Two potential examples uncovered during this examination are the p75 neurotrophin receptor [extracellular domain] and diacetylspermine (a polyamine). See Table 1 and the Supplementary materials for further discussion.

### 6.5. Indirect markers of health/wellness: dietary markers

An alternative approach to endogenous biomarkers for gauging health (wellness) would target metabolic products of exogenous health-protective or health-promoting chemicals, namely those that are associated with healthy nutrition. These metabolites would be selected to indicate exposure (via oral consumption) to beneficial, "functional" foods—those foods whose intake is strongly associated with improved health; the converse approach would be to target metabolites of those exogenous dietary chemicals that are components of harmful foods or which are naturally occurring toxic xenobiotics associated with foods. Foods comprise tens

of thousands of chemical constituents (Scalbert et al., 2014). Most are endogenous constituents of foods, while others are exogenous, natural product contaminants such as the ubiquitous mycotoxins (e.g., see recent symposium: Jackson and Ryu, 2017). The mercapturic acid conjugates (*N*-acetylcysteine *S*-conjugates) of various thiocyanates are examples that originate from their glucosinolate precursors naturally present in dietary cruciferous vegetables (e.g., Vermeulen et al., 2006); a specific example is the *N*-acetylcysteine *S*-conjugate of sulforaphane, which itself is a metabolite of the precursor glucoraphanin. A major criterion is that these markers must be endogenous metabolites of parent chemicals unique to these foods; they should also not be widely used in commercial nutritional supplements. This ensures that their occurrence in sewage reflects actual dietary consumption rather than the introduction/disposal of raw/cooked foods or dietary supplements to sewers; and of course, they must also be stable in sewage.

Dietary biomarkers are often classified as representing short-, intermediate-, and long-term consumption (a function of their half-lives and excretion kinetics). But this factor does not matter for BioSCIM because it would continually integrate the collective exposures for members of a population.

Examples of candidate markers for consumption of nutritious foods include the alkylresorcinols, which derive from the phenolic lipids in the bran fraction of whole grain wheat and rye, and which are absent from refined grains (Ross, 2012). Specific examples are 3,5-dihydroxybenzoic acid (DHBA) and 3-(3,5-dihydroxyphenyl)-propanoic acid (DHPPA). Another example is tyrosol, a metabolite of oleuropein, which is a secoiridoid responsible for the bitter, burning sensation from extra-virgin olive oil (Piroddi et al., 2016); note that hydroxytyrosol (another secoiridoid in olive oil) would not suffice as a biomarker because it can also originate as a metabolic product from dopamine oxidation, where it is referred to as DOPET (3,4-dihydroxyphenylethanol). The state of research on dietary markers undergoes continual review (e.g., see: Hedrick et al., 2012; Scalbert et al., 2014).

The polyphenols are another example of chemicals presumed to be associated with nutritious foods. Some candidates as indicator metabolites that could reflect consumption include: dihydroferulic acid (for coffee), gallic acid ethyl ester (for red wine), naringenin glucuronide (for citrus fruit), 4-*O*-methylgallic acid (for tea), phloretin glucuronide (for apples and pears), and *O*-methyl epicatechin sulfate (for chocolate) (Edmands et al., 2015).

Examples of candidate markers for consumption of potentially harmful foods include: metabolites of ethyl alcohol [ethyl glucuronide (EtG), ethyl sulfate (EtS), phosphatidyl ethanol (PEth)] and 5-hydroxytryptophol (HTOL) (Maenhout et al., 2013; Reid et al., 2011) (note that since PEth is synthesized via phospholipase D, it serves as a definitive metabolite, but one for which sewage has apparently not yet been monitored); DON-15-GlcA, which is the most important glucuronic acid (GlcA) conjugate of deoxynivalenol (DON; also known as vomitoxin)—one of the most prevalent *Fusarium* mycotoxins (common contaminant of grains worldwide, especially wheat, barley, maize, and oats) (Gruber-Dorninger et al., 2017; Huybrechts et al., 2015); and 1-

methyl-histidine (a marker for meat consumption), but not 3-methyl-histidine, which can be formed via muscle catabolism (Fraser et al., 2016; Lindsay and Costello, 2017). Note, however, that although conjugates provide an assurance that they originate from metabolism and therefore reflect exposure, their inherent chemical instability introduces added challenges in performing back-calculations to quantify exposure (e.g., Banks et al., 2017).

## 7. Biomarker candidates meriting evaluation for use with BioSCIM

One of the major objectives of this paper was to examine the published literature for additional endogenous human biomarkers to consider as potential targets for BioSCIM. During this work, over 400 articles were selected from a wide array of journals and books spanning all fields involved with biomarker research (e.g., metabolomics, clinical diagnostics, and sports medicine research). These were entered into a bibliographic database (EndNote X8, Clarivate Analytics) and examined for their potential relevance. Both forward and backward citation searching were used to augment and guide keyword searches. Ideally, the ultimate objective would be to compile a suite of biomarkers that could be used as BioSCIM targets and which represented a spectrum of different disease processes—analogous to a conventional clinical urinalysis panel. These candidate biomarkers are summarized in Table 1. More detailed background narrative is presented for each biomarker in the Supplementary material. These select few markers certainly do not compose an exhaustive list. The intention was instead to offer some examples that illustrate the type of information required to assess whether a biomarker might serve as a suitable candidate and therefore deserves more in-depth, future evaluation.

These new candidate biomarkers must meet the same criteria that were originally defined for the archetype biomarker—the isoprostane class (Daughton, 2012b). Biomarkers suitable for BioSCIM must meet a number of criteria, many of which never require consideration for use in clinical medicine. The most important additional criteria (summarized from earlier discussion) are:

(1) Origination from endogenous human metabolism (and notably, not being produced de novo by microbial activity in sewage or biofilms);

(2) Extensive urinary excretion (fecal excretion is more problematic for monitoring);

(3) Minimal additional origins from raw, cooked, or digested foods (all of which could confound monitoring data);

(4) Sufficient molecular stability in sewage (resisting chemical alteration by physicochemical processes—such as via photolysis—and by microbial transformation);

(5) Excreted in quantities sufficiently high to allow for analytical detection after dilution in sewage and despite its matrix being more complex than urine;

(6) Excreted quantities are sufficiently sensitive to changes in health status (i.e., detectable changes in the resulting signal:noise ratio);

(7) Excreted levels trend in a consistent direction as a function of escalating or declining stressor exposure (e.g., absence of inverted-U or J-shaped dose-response or mixed monotonic dose-response; see example of HPMA in the section) [see section: Biomarkers of positive health versus disease]; and

(8) Excreted quantities are minimally influenced by drug therapy or diet.

Criteria 3–5 add considerable difficulty in assessing the published literature, as this information is generally not available in clinical publications on biomarkers. Instead, searches for this information need to be extended to fields such as nutrition, natural products, and microbial processes. De novo biosynthesis in sewage is another aspect that can

be difficult to ascertain or predict. Of these concerns, the difficulty in locating sufficient data regarding in situ biotransformation in sewage shows that Criterion 4 could be anticipated as posing a major challenge for finding endogenous biomarkers suitable for BioSCIM (e.g., see: McCall et al., 2016; Ramin et al., 2016). With respect to monitoring sewage for endogenous biomarkers, molecular stability poses specific questions with regard to suitable endogenous biomarkers that happen to be proteins or peptides. An obvious question is whether they can survive structural degradation or transformation during transit to the WWTP (primarily via microbial proteolytic processes). Would these BioSCIM biomarkers have sufficient persistence in sewage? Note that monitoring for the purposes of BioSCIM would focus only on the collection/distribution system of raw sewage and on the influent stream to WWTPs. Because treated sewage effluent would not play a role in monitoring, the targeted biomarkers would not experience the presumably most rigorous conditions for degradation.

Little is known regarding the stability or persistence of specific proteins in sewage. But one class unrelated to potential BioSCIM markers could possibly serve to inform the discussion. The proteinaceous biopharmaceuticals (sometimes called biologics) are generally assumed to have short half-lives and not survive sewage treatment unaltered. Historically, they have therefore not warranted regulatory environmental assessments as pharmaceuticals (e.g., Kümmerer, 2009; Straub, 2016). Some prime examples are the monoclonal antibody biologics (e.g., the mab's such as infliximab, rituximab, trastuzumab, abciximab, adalimumab). But even with these proteins, little has been published.

With this said, there are still two sides to the debate regarding protein stability in sewage, especially given that a select few proteins are known to have exceptional stability in the environment (the prions being an archetype). While proteins are quite stable to non-enzymatic degradation, they are usually ready substrates for proteases. But proteolysis by extracellular proteases could be expected to vary dramatically depending on the types of proteins, the types and levels of specific proteases present, and the very low expected levels of BioSCIM biomarkers compared with the preferred substrates, which may be present at far higher levels (resulting in competitive inhibition). Overall, little is known about the fate of specific proteins in sewage, as there has been little incentive for its general study (Westgate, 2009; Westgate and Park, 2010).

With respect to endogenous biomarkers targeted in BioSCIM, however, some could be considered more closely related to non-ribosomal peptides (NRPs), for which much less is known regarding their breakdown than ribosomal proteins. Indeed, many NRPs are commercial pharmacologic agents, such as antibiotics (e.g., bacitracin), cytostatics (e.g., bleomycin), and immunosuppressants (cyclosporine). Many of these NRPs have been reported in sewage or streams receiving treated sewage—evidence that they at least partly survive breakdown. A host of other NRPs are natural product toxins, generally from the secondary metabolism of microorganisms, such as the microcystins. NPRs are much more resistant to proteolysis than human ribosomal proteins (Lundeen et al., 2016). Any assessment of the projected half-lives of BioSCIM biomarkers with peptide bonds would therefore probably entail the empirical study of each specific biomarker.

The following study serves as an illustrative instance relevant to the third criterion (above), where the possibility of dietary sources for a biomarker must be assessed. It is rather remarkable that dietary sources are usually not considered as possible confounding sources for endogenous production (as they can bias clinical urinary data) and also as having biological activity of their own. Symmetric dimethylarginine (SDMA) originates from the post-translational methylation of proteins (not from free arginine). SDMA has received substantial attention over the years as a direct indicator of renal dysfunction. It is slowly metabolized, and as a result it is extensively excreted in urine (in contrast to its isomer, asymmetric dimethylarginine). It is also biologically active itself, playing a role, for example, in the development of cardiovascular disease (Schepers et al., 2014). With this said, it was only in 2013 that

SDMA (as well as the other two methylated arginines) was found to occur naturally in a broad spectrum of common foods (Servillo et al., 2013). The importance of this with respect to BioSCIM, regardless of whether diet serves as a systemic source of SDMA, is that dietary SDMA could bias excreted levels resulting from endogenous production. SDMA might have otherwise made a good candidate for BioSCIM.

There are numerous challenges faced in searching the published literature for possible candidate biomarkers. The sheer size of the literature can be daunting. It is also populated with repetitive articles for any given biomarker that add but limited, incremental knowledge. And very few articles prove useful for evaluating the criteria summarized above, as they are only peripherally relevant. No claim is made that the candidate biomarkers presented here would be ideal for BioSCIM. Instead these are being presented with the understanding that they represent a concerted effort designed for others to build upon and to catalyze research designed to evaluate their utility with BioSCIM. Undoubtedly, additional, new biomarkers useful for BioSCIM will eventually emerge from the efforts of others; for example, refer to an overview by Gao (2015a).

Note that most of the biomarkers presented here (Table 1 and Supplementary material) serve as measures for stress, injury, or disease. Very few are potential candidates for measuring health, wellness, or eustress. Most of these markers serve as integrative measures for multiple processes. Very few reflect specific diseases. And finally, to reiterate, it is critical for successful application of BioSCIM to have at least one biomarker of health for the implementation of the "disease:health biomarker ratio" concept—d:hBR [see section: New conceptual approach for BioSCIM].

## 8. The future

Beyond the articles cited at the beginning of this paper (those published since 2012), the prospects of monitoring sewage for endogenous biomarkers for gauging the status of community-wide health has begun to attract additional interest with respect to smart and sustainable cities (e.g., Poletti and Treville, 2016)—one example being the "Underworlds" project at MIT's Senseable City Lab (Fitzgerald, 2015; Graber, 2017; Reis-Castro, 2017). It has also become a focus of transdisciplinary research under the European COoperation in Scientific and Technology (COST) program (COST, 2013). The European COST program is a natural extension of the research conducted primarily in Europe since 2005 on the measurement of illicit drugs in sewage to gauge community-wide consumption (i.e., "sewage epidemiology"). The following are some additional points regarding the future of BioSCIM.

### 8.1. Personalized community health

The ultimate value or power of BioSCIM in serving as a gauge for collective community-wide health will be a function primarily of the numbers of endogenous biomarkers (and to a lesser degree, exogenous markers) that can be reliably monitored in sewage. Given a suite of orthogonal biomarkers with sufficient diagnostic or prognostic power, BioSCIM might serve as a major tool simply for alerting and motivating communities and individuals to the need for design and implementation of interventions that promote healthy behaviors tailored to their geographic locales. BioSCIM could eventually serve as a tool for the integration of medical-based monitoring approaches such as P4H (Sagner et al., 2016) into everyday living—even if solely for improving a population's overall health literacy and to promote health vigilance. Other examples of existing programs with which BioSCIM could interact or inform include the *Healthy People 2020* initiative and the large-scale health assessment program "Community Health Status Indicators" (CHSI), a program managed by the Centers for Disease Control and Prevention (CDC, 2013, 2015). The CHSI program makes use of an assortment of conventional health indicators (and health determinants) that have been compiled from a variety of published sources (e.g., see:

CDC, 2015). Significantly, the CDC has placed the emphasis on "improving community health" (e.g., via informing decision makers) rather than on monitoring or assessing it. While BioSCIM could eventually serve as another tool for programs such as CHSI, these types of programs could reciprocally serve as a means for benchmarking BioSCIM—by comparing sewage-biomarker data with existing health assessments for particular communities. In fact, BioSCIM might be viewed as the means for creating a new perspective for medicine—"personalized community health"—with the objective being to enhance the extent of community well-being. Of course, a major question not yet addressed is what exactly is meant by "collective health" of a population?

### 8.2. Concerns potentially regarding ethics and mass surveillance

Even should BioSCIM eventually succeed in meeting its intended objectives, attention will need to be devoted to strategies for addressing public concerns regarding its potential for misuse or abuse. These concerns (whether real or imagined) surround what some fear could be unethical applications, in particular surveillance that could reveal personal information. This concern has already been encountered with the application of sewage epidemiology for monitoring drugs of abuse and illicit drugs (e.g., see: Castiglioni, 2016; Hall et al., 2012; Hering, 2009; Prichard et al., 2014; Prichard et al., 2015).

For BioSCIM to be useful in advancing community health, the overall approach must be embraced by the public and by municipalities. And concerns extend far beyond those of individual privacy. With sewage epidemiology and illicit drugs, one form of resistance has been that some cities opt out of monitoring because they do not want to be perceived as "hotspots" for illicit drug use—whether true or not. Similarly, with BioSCIM, perhaps some cities or communities would want to avoid notoriety for being labeled as "unhealthy", especially if being compared against a nearby city or adjacent community. There are many potential ramifications, not the least of which could be adverse effects on the well-being of citizens, impacts on property values, and adjustment of insurance premiums by health insurers. What might happen with a community that appears to have impaired collective health when the responsible stressor(s) cannot be identified or cost-effective interventions cannot be implemented? With the advent of high-resolution street-level air pollution mapping in near-real time, many of these same concerns have already been captured and articulated by Apte et al. (2017): "Broader societal consequences of the public awareness enabled by high-resolution pollution maps might include shifts in urban land-use decisions, regulatory actions, and in the political economy of environmental 'riskscapes'."

These potential concerns reveal the opportunities for involvement of other disciplines in the early development of BioSCIM—obvious examples being social psychologists, science communicators, and ethicists. Even if BioSCIM were eventually shown to be a valuable tool for quickly and inexpensively monitoring community health, any widespread deployment could be thwarted by public misperceptions or by inadequate controls on unethical use; two of many possible examples might be the surreptitious monitoring of a community's sewage for the purpose of depressing its property values or to show that local industrial point sources of known hazardous pollutants are not correlated with biomarkers of disease.

### 8.3. BioSCIM and the exposome

BioSCIM could serve as the first tool for the real-world implementation at the community level of the *exposome* concept (Wild, 2005). The intention would be to shift the primary focus from exposure to bona fide health outcomes. BioSCIM could be viewed as the first way to begin examining the exposome in its most meaningful and ultimate state—namely, the context of the real world at the population level. BioSCIM can be viewed as the application of exposomics at the population level rather than at the level of the individual. By targeting an

appropriate suite of orthogonal, endogenous human biomarkers, BioSCIM would serve as an integrative measure of exposure for all stressors at play (chemical and non-chemical alike). BioSCIM could also be applied to specific sub-populations. Some examples include schools, hospitals, military installations, prisons, and occupational settings. BioSCIM would hold another advantage over the use of epidemiological cohorts, which often suffer from the vulnerability of selection bias, such as under-representation by certain groups, especially those suffering from socioeconomic or health disparities (Wild, 2012).

With this said, it is important to recognize a potential major vulnerability of BioSCIM. By necessarily measuring the "average" collective levels of biomarkers for a community, the contributions from the extremes are obscured; the relative importance of the extremes would also be obscured. This could limit the utility of BioSCIM in identifying the existence of outliers that might represent particularly vulnerable sub-populations or those that have received unusually high exposures (Gochfeld and Burger, 2011). After all, sometimes a particular type and level of exposure might be benign and other times the very same exposure could be adverse. The ultimate exposure outcome is a function of the context of the exposure, which is dictated by an exceedingly complex interplay of myriad factors—many of which have yet to be uncovered.

As a tool to study the exposome at the community scale, a specific community would have to show relatively stable time-course levels for each biomarker. With sufficiently stable short-term levels, trends or changes in health within a given community could be revealed. For example, community-wide levels of a biomarker that begin to trend away from the preexisting stable level could indicate: (1) community-wide exposure to a newly present emerging stressor(s) or (2) increasing levels of an existing stressor within the community. A trend in the opposite direction could indicate: (1) the diminution of a long-present stressor or (2) the emergence of a protective factor (negating the action of a long-present stressor). BioSCIM's future is currently a function of the imagination and innovation of those who become engaged with its development and implementation.

### 8.4. The urinary proteome

Historically, urinary protein biomarkers have received far less attention that those in blood (Gao, 2015a), probably because the overall protein content of urine from healthy individuals is low (less than 150 mg/day); ironically though, the far higher protein content of serum is dominated by a few very abundant proteins, which makes the ambient background in blood more challenging for analysis. As chemical analysis tools have improved, it has become evident that the types of proteins in urine reflect those in blood. After all, much of the urinary protein originates from the glomerular filtration of blood, although a majority originates from direct production by the kidney and urogenital tract (Gao, 2015a). Note, however, that renal function (glomerular filtration rates) can change as a consequence of toxicity from the very same stressors that might partly participate in the overall stress that BioSCIM intends to measure. This gets entangled with the need to normalize excreted levels (Weaver et al., 2016).

Efforts to identify the growing numbers of proteins that were being reported in urine began in earnest in the late 1970s (e.g., see: Anderson et al., 1979). Research aimed at revealing and profiling the urinary proteome would not begin until the 1990s when the limitations of chromatography, mass spectrometry, and high-speed data acquisition/analysis began to rapidly fall away (e.g., Spahr et al., 2001). The first large-scale proteomic analysis of urine was reported only in 2006 and was made possible by the application of high-resolution mass spectrometry (Adachi et al., 2006); subsequent major advances include those of others such as Marimuthu et al. (2011). The identification of new urinary biomarkers will continue to escalate. And the ease and speed of analysis will continually improve. The utility of urinary exosomes (especially with regard to other classes of biomarkers such as mRNA and

miRNA) could dramatically expand the number of useful urinary biomarkers. These developments could accelerate the development and implementation of BioSCIM applications. BioSCIM could eventually be viewed as an endeavor that involves monitoring the sewage human metabolome.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2017.11.102.

### EPA notice

### References

Adachi, J., Kumar, C., Zhang, Y., Olsen, J.V., Mann, M., 2006. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. Genome Biol. 7, R80.

Alexander, G.C., Stevens, R.J., 1976. Per capita phosphorus loading from domestic sewage. Water Res. 10, 757–764.

Anderson, N.G., Anderson, N.L, Tollaksen, S.L., 1979. Proteins of human urine. I. Concentration and analysis by two-dimensional electrophoresis. Clin. Chem. 25, 1199–1210.

Apte, J.S., Messier, K.P., Gani, S., Brauer, M., Kirchstetter, T.W., Lunden, M.M., et al., 2017. High-resolution air pollution mapping with Google street view cars: exploiting big data. Environ. Sci. Technol. 51, 6999–7008.

Arnold, C., 2016. Pipe dreams: tapping into the health information in our sewers. Environ. Health Perspect. 124, A86–A91.

Avioli, L.V., Lee, S.W., McDonald, J.E., Lund, J., DeLuca, H.F., 1967. Metabolism of vitamin $D_3$-$^3H$ in human subjects: distribution in blood, bile, feces, and urine. J. Clin. Investig. 46, 983–992.

Baker, D.R., Barron, L., Kasprzyk-Hordern, B., 2014. Illicit and pharmaceutical drug consumption estimated via wastewater analysis. Part A: chemical analysis and drug use estimates. Sci. Total Environ. 487, 629–641.

Banks, A.P.W., Lai, F.Y., Mueller, J.F., Jiang, G., Carter, S., Thai, P.K., 2017. Potential impact of the sewer system on the applicability of alcohol and tobacco biomarkers in wastewater-based epidemiology. Drug Test. Anal.:1–9 https://doi.org/10.1002/dta.2246.

Bartoszewicz, Z., Kondracka, A., Jazwiec, R., Popow, M., Dadlez, M., Bednarczuk, T., 2013. Can we accurately measure the concentration of clinically relevant vitamin D metabolites in the circulation? The problems and their consequences. Endokrynol. Pol. 64, 238–245.

Boutin, C., Eme, C., 2016. Domestic wastewater characterization by emission source. 13th IWA Specialized Conference on Small Water and Wastewater Systems, 14–16 September. International Water Association, Athens, Greece 8 pp. http://uest.ntua.gr/swws/proceedings/presentation/06.Boutin.pdf https://hal.archives-ouvertes.fr/hal-01469077/document (accessed 13 July 2017).

Brown, L., Rhead, M.M., Braven, J., 1984. Initial studies on the use of high-performance liquid chromatography for the rapid assessment of sewage treatment efficiency. Water Res. 18, 955–961.

Brown, L., Braven, J., Rhead, M.M., 1985. Method and Apparatus for Monitoring Sewage Treatment Efficiency and Determining Sewage Sources. Patent No: 4,554,077. Issue Date: November 19 1985; Available on:. https://www.google.com/patents/US4554077, Accessed date: 13 July 2017.

Bruno, R., Hall, W., Kirkbride, K.P., Lai, F.Y., O'Brien, J.W., Prichard, J., et al., 2014. Commentary on Ort et al. (2014): what next to deliver on the promise of large scale sewage-based drug epidemiology? Addiction 109, 1353–1354.

Bugajski, P., Chmielowski, K., Kaczor, G., 2016. Optimizing the percentage of sewage from septic tanks for stable operation of a wastewater treatment plant. Pol. J. Environ. Stud. 25, 1421–1425.

Castiglioni S, (ed.), Assessing Illicit Drugs in Wastewater: Advances in Wastewater-based Drug Epidemiology. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), Luxembourg, 2016, Volume Insights 22, 82 pp.; doi:https://doi.org/10.2810/017397. (accessed 13 July 2017).

Castiglioni, S., Bijlsma, L., Covaci, A., Emke, E., Hernández, F., Reid, M., et al., 2013. Evaluation of uncertainties associated with the determination of community drug use through the measurement of sewage drug biomarkers. Environ. Sci. Technol. 47, 1452–1460.

CDC, 2013. Community Health Assessment for Population Health Improvement: Resource of Most Frequently Recommended Health Outcomes and Determinants. U.S. Centers for Disease Control and Prevention, Office of Surveillance, Epidemiology, and Laboratory Services, Atlanta, GA:p. 29 Available on:. http://wwwn.cdc.gov/CommunityHealth/PDF/Final_CHAforPHI_508.pdf, Accessed date: 13 July 2017.

CDC, 2015. Community Health Status Indicators (CHSI 2015), Last Updated 13 January 2015. Centers for Disease Control and Prevention, U.S. Department of Health & Human Services, Atlanta, GA Available on:. https://wwwn.cdc.gov/CommunityHealth/HomePage.aspx, Accessed date: 13 July 2017.

Chen, C., Kostakis, C., Gerber, J.P., Tscharke, B.J., Irvine, R.J., White, J.M., 2014. Towards finding a population biomarker for wastewater epidemiology studies. Sci. Total Environ. 487, 621–628.

COST, 2013. Memorandum of Understanding for the Implementation of a European Concerted Research Action Designated as COST Action ES1307: Sewage Biomarker Analysis for Community Health Assessment 2013, COST 054/13 (22 November):25. European Cooperation in Scientific and Technology (COST). Available on:. http://w3.cost.eu/fileadmin/domain_files/ESSEM/Action_ES1307/mou/ES1307-e.pdf www.cost.eu/download/39268 (accessed 13 July 2017).

Daughton, C.G., 2011. Illicit drugs: contaminants in the environment and utility in forensic epidemiology. Rev. Environ. Contam. Toxicol. 210, 59–110.

Daughton, C.G., 2012a. Real-time estimation of small-area populations with human biomarkers in sewage. Sci. Total Environ. 414, 6–21.

Daughton, C.G., 2012b. Using biomarkers in sewage to monitor community-wide human health: Isoprostanes as conceptual prototype. Sci. Total Environ. 424, 16–38.

Eckert, E., Schmid, K., Schaller, B., Hiddemann-Koca, K., Drexler, H., Göen, T., 2011. Mercapturic acids as metabolites of alkylating substances in urine samples of German inhabitants. Int. J. Hyg. Environ. Health 214, 196–204.

Edmands, W.M., Ferrari, P., Rothwell, J.A., Rinaldi, S., Slimani, N., Barupal, D.K., et al., 2015. Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across European countries. Am. J. Clin. Nutr. 102, 905–913.

Eggimann, S., Mutzner, L., Wani, O., Schneider, M.Y., Spuhler, D., Moy de Vitry, M., et al., 2017. The potential of knowing more: a review of data-driven urban water management. Environ. Sci. Technol. 51, 2538–2553.

Fielding, J.E., Kumanyika, S., Manderscheid, R.W., 2013. A perspective on the development of the Healthy People 2020 framework for improving US population health. Public Health Rev. 35, 1–24.

Fitzgerald, M., 2015. What Does Cambridge Sewage Say About Residents? MIT Plans to Find Out. A New Project Aims to Understand a City's Health And Habits Through Its Waste. Boston Globe Media Partners, LLC, Boston, MA Available on:. http://www.bostonglobe.com/ideas/2015/01/09/what-does-cambridge-sewage-say-about-residents-mit-plans-find-out/qxBK9jbHYJFMcNGiTM2L2I/story.html, Accessed date: 13 July 2017.

Fortin, M.-C., Carrier, G., Bouchard, M., 2008. Concentrations versus amounts of biomarkers in urine: a comparison of approaches to assess pyrethroid exposure. Environ. Health 7, 55.

Fox, M.A., Brewer, L.E., Martin, L., 2017. An overview of literature topics related to current concepts, methods, tools, and applications for cumulative risk assessment (2007–2016). Int. J. Environ. Res. Public Health 14, 389.

Fraser, G.E., Jaceldo-Siegl, K., Henning, S.M., Fan, J., Knutsen, S.F., Haddad, E.H., et al., 2016. Biomarkers of dietary intake are correlated with corresponding measures from repeated dietary recalls and food-frequency questionnaires in the Adventist Health Study-2. J. Nutr. 146, 586–594.

Friedler, E., Butler, D., Alfiya, Y., 2013. Wastewater composition. In: Larsen, T.A., Udert, K.M., Lienert, J. (Eds.), Source Separation and Decentralization for Wastewater Management. IWA Publishing, London, UK, pp. 241–257 (chapter 17).

Füzéry, A.K., Levin, J., Chan, M.M., Chan, D.W., 2013. Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. Clin. Proteomics 10, 13.

Galano, J.M., Lee, Y.Y., Oger, C., Vigor, C., Vercauteren, J., Durand, T., et al., 2017. Isoprostanes, neuroprostanes and phytoprostanes: an overview of 25 years of research in chemistry and biology. Prog. Lipid Res. 68, 83–108.

Gao, Y. (Ed.), 2015a. Urine Proteomics in Kidney Disease Biomarker Discovery. Advances in Experimental Medicine and Biology 845. Springer Netherlands, Dordrecht 203 pp. http://link.springer.com/book/10.1007%2F978-94-017-9523-4, Accessed date: 13 July 2017.

Gao, Y., 2015b. Urine is a better biomarker source than blood especially for kidney diseases. In: Gao, Y. (Ed.), Urine Proteomics in Kidney Disease Biomarker Discovery, Advances in Experimental Medicine and Biology 845. Springer Netherlands, Dordrecht, pp. 3–12.

Gao, J., O'Brien, J., Du, P., Li, X., Ort, C., Mueller, J.F., et al., 2016. Measuring selected PPCPs in wastewater to estimate the population in different cities in China. Sci. Total Environ. 568, 164–170.

Gao, J., Banks, A., Li, J., Jiang, G., Lai, F.Y., Mueller, J.F., et al., 2017. Evaluation of in-sewer transformation of selected illicit drugs and pharmaceutical biomarkers. Sci. Total Environ. 609, 1172–1181.

Gavina, J.M.A., Yao, C., Feng, Y.-L., 2014. Recent developments in DNA adduct analysis by mass spectrometry: a tool for exposure biomonitoring and identification of hazard for environmental pollutants. Talanta 130, 475–494.

Gaw, S., Glover, C.N., 2016. A case of contagious toxicity? Isoprostanes as potential emerging contaminants of concern. Sci. Total Environ. 560–561, 295–298.

Ghini, V., Saccenti, E., Tenori, L., Assfalg, M., Luchinat, C., 2015. Allostasis and resilience of the human individual metabolic phenotype. J. Proteome Res. 14, 2951–2962.

Gochfeld, M., Burger, J., 2011. Disproportionate exposures in environmental justice and other populations: the importance of outliers. Am. J. Public Health 101, S53–S63.

González-Mariño, I., Rodil, R., Barrio, I., Cela, R., Quintana, J.B., 2017. Wastewater-based epidemiology as a new tool for estimating population exposure to phthalate plasticizers. Environ. Sci. Technol. 51, 3902–3910.

Graber, C., 2017 January 19. Why an MIT Robot is Collecting Poop From Our Sewers. The Boston Globe Magazine, Boston, MA Available on:. https://www.bostonglobe.com/magazine/2017/01/19/why-mit-has-robot-collecting-samples-sewer/g1EhuRPbVR2VtwwJuo2DTN/story.html, Accessed date: 13 July 2017.

Gracia-Lor, E., Castiglioni, S., Bade, R., Been, F., Castrignanò, E., Covaci, A., et al., 2017. Measuring biomarkers in wastewater as a new source of epidemiological information: current state and future perspectives. Environ. Int. 99, 131–150.

Gruber-Dorninger, C., Novak, B., Nagl, V., Berthiller, F., 2017. Emerging mycotoxins: beyond traditionally determined food contaminants. J. Agric. Food Chem. 65, 7052–7070.

Gruys, E., Toussaint, M.J.M., Niewold, T.A., Koopmans, S.J., 2005. Acute phase reaction and acute phase proteins. J Zhejiang Univ Sci B 6, 1045–1056.

Hall, W., Prichard, J., Kirkbride, P., Bruno, R., Thai, P.K., Gartner, C., et al., 2012. An analysis of ethical issues in using wastewater analysis to monitor illicit drug use. Addiction 107, 1767–1773.

Hedrick, V.E., Dietrich, A.M., Estabrooks, P.A., Savla, J., Serrano, E., Davy, B.M., 2012. Dietary biomarkers: advances, limitations and future directions. Nutr. J. 11, 1–14.

Hering, C.L., 2009. Flushing the fourth amendment down the toilet: how community urinalysis threatens individual privacy. Ariz. Law Rev. 53, 741–776.

Hildonen, S., Skarpen, E., Halvorsen, T.G., Reubsaet, L., 2016. Isolation and mass spectrometry analysis of urinary extraexosomal proteins. Sci. Rep. 6, 36331.

Hocher, B., Adamski, J., 2017. Metabolomics for clinical use and research in chronic kidney disease. Nat. Rev. Nephrol. 13, 269–284.

Holick, M.F., 2011. Vitamin D: evolutionary, physiological and health perspectives. Curr. Drug Targets 12, 4–18.

Horst, R.L., Reinhardt, T.A., Reddy, G.S., 2005. Vitamin D metabolism. In: Feldman, D., Pike, J.W., Glorieux, F.H. (Eds.), Vitamin D. 1. Elsevier Academic Press, London, UK, pp. 15–36.

Huebner, A.R., Somparn, P., Benjachat, T., Leelahavanichkul, A., Avihingsanon, Y., Fenton, R.A., et al., 2015. Exosomes in urine biomarker discovery. In: Gao, Y. (Ed.), Urine Proteomics in Kidney Disease Biomarker Discovery. Springer Netherlands, Dordrecht, pp. 43–58.

Huybrechts, B., Martins, J.C., Debongnie, P., Uhlig, S., Callebaut, A., 2015. Fast and sensitive LC–MS/MS method measuring human mycotoxin exposure using biomarkers in urine. Arch. Toxicol. 89, 1993–2005.

Jackson, L.S., Ryu, D., 2017. Summary of the ACS symposium on public health perspectives of mycotoxins in food. J. Agric. Food Chem. 65, 7017–7020.

Jain, K.K., 2010. The Handbook of Biomarkers. Humana Press, Springer, New York 492 pp. http://www.springer.com/us/book/9781607616849, Accessed date: 13 July 2017.

Jaramillo, R., Shuck, S.C., Chan, Y.S., Liu, X., Bates, S.E., Lim, P.P., et al., 2017. DNA Advanced Glycation End Products (DNA-AGEs) are elevated in urine and tissue in an animal model of type 2 diabetes. Chem. Res. Toxicol. 30, 689–698.

Javeed, N., Mukhopadhyay, D., 2016. Exosomes and their role in the micro-/macro-environment: a comprehensive review. J. Biomed. Res. 30, 1–9.

Jones, H.E., Hickman, M., Kasprzyk-Hordern, B., Welton, N.J., Baker, D.R., Ades, A.E., 2014. Illicit and pharmaceutical drug consumption estimated via wastewater analysis. Part B: placing back-calculations in a formal statistical framework. Sci. Total Environ. 487, 642–650.

Kookana, R.S., Williams, M., Boxall, A.B.A., Larsson, D.G.J., Gaw, S., Choi, K., et al., 2014. Potential ecological footprints of active pharmaceutical ingredients: an examination of risk factors in low-, middle- and high-income countries. Philos. Trans. R. Soc. B Biol. Sci. 369, 20130586.

Kümmerer, K., 2009. The presence of pharmaceuticals in the environment due to human use—present knowledge and future challenges. J. Environ. Manag. 90, 2354–2366.

Lai, F.Y., Anuj, S., Bruno, R., Carter, S., Gartner, C., Hall, W., et al., 2015. Systematic and day-to-day effects of chemical-derived population estimates on wastewater-based epidemiology. Environ. Sci. Technol. 49, 999–1008.

Lamichhane, K.M., Babcock Jr., R.W., 2013. Survey of attitudes and perceptions of urine-diverting toilets and human waste recycling in Hawaii. Sci. Total Environ. 443, 749–756.

Landry, K.A., Boyer, T.H., 2016. Life cycle assessment and costing of urine source separation: focus on nonsteroidal anti-inflammatory drug removal. Water Res. 105, 487–495.

Ledger, J.E., Watson, G.J., Ainley, C.C., Compston, J.E., 1985. Biliary excretion of radioactivity after intravenous administration of $^3$H-1,25-dihydroxyvitamin $D_3$ in man. Gut 26, 1240–1245.

Lindsay, A., Costello, J.T., 2017. Realising the potential of urine and saliva as diagnostic tools in sport and exercise medicine. Sports Med. 47, 11–31.

Logue, C., Dowey, L.R.C., Strain, J.J., Verhagen, H., McClean, S., Gallagher, A.M., 2017. Application of liquid chromatography-tandem mass spectrometry to determine urinary concentrations of five commonly used low-calorie sweeteners: a novel biomarker approach for assessing recent intakes? J. Agric. Food Chem. 65, 4516–4525.

Lopardo, L., Cummins, A., Rydevik, A., Kasprzyk-Hordern, B., 2017. New analytical framework for verification of biomarkers of exposure to chemicals combining human biomonitoring and water fingerprinting. Anal. Chem. 89, 7232–7239.

Lundeen, R.A., Chu, C., Sander, M., McNeill, K., 2016. Photooxidation of the antimicrobial, nonribosomal peptide bacitracin a by singlet oxygen under environmentally relevant conditions. Environ. Sci. Technol. 50, 8586–8595.

Maenhout, T.M., De Buyzere, M.L., Delanghe, J.R., 2013. Non-oxidative ethanol metabolites as a measure of alcohol intake. Clin. Chim. Acta 415, 322–329.

Marimuthu, A., O'Meally, R.N., Chaerkady, R., Subbannayya, Y., Nanjappa, V., Kumar, P., et al., 2011. A comprehensive map of the human urinary proteome. J. Proteome Res. 10, 2734–2743.

Martinez-Subiela, S., Ceron, J., 2005. Evaluation of acute phase protein indexes in dogs with leishmaniasis at diagnosis, during and after short-term treatment. Vet. Med. 50, 39–46.

McCall, A.-K., Bade, R., Kinyua, J., Lai, F.Y., Thai, P.K., Covaci, A., et al., 2016. Critical review on the stability of illicit drugs in sewers and wastewater samples. Water Res. 88, 933–947.

McCall, A.-K., Palmitessa, R., Blumensaat, F., Morgenroth, E., Ort, C., 2017. Modeling in-sewer transformations at catchment scale—implications on drug consumption estimates in wastewater-based epidemiology. Water Res. 122, 655–668.

Milne, G.L., 2017. Classifying oxidative stress by $F_2$-isoprostane levels in human disease: the re-imagining of a biomarker. Redox Biol. 12, 897–898.

Mitchell, P.J., Welton, J., Staffurth, J., Court, J., Mason, M.D., Tabi, Z., et al., 2009. Can urinary exosomes act as treatment response markers in prostate cancer? J. Transl. Med. 7, 1–13.

Mora, E.M., Álvarez-Cubela, S., Oltra, E., 2016. Biobanking of exosomes in the era of precision medicine: are we there yet? Int. J. Mol. Sci. 17, 13.

Nagarajah, S., 2016. Exosome secretion—more than simple waste disposal? Implications for physiology, diagnostics and therapeutics. J. Circ. Biomark. 5, 7.

Nakada, N., Hanamoto, S., Jürgens, M.D., Johnson, A.C., Bowes, M.J., Tanaka, H., 2017. Assessing the population equivalent and performance of wastewater treatment through the ratios of pharmaceuticals and personal care products present in a river basin: application to the River Thames basin, UK. Sci. Total Environ. 575, 1100–1108.

O'Brien, J.W., Thai, P.K., Eaglesham, G., Ort, C., Scheidegger, A., Carter, S., et al., 2014. A model to estimate the population contributing to the wastewater using samples collected on census day. Environ. Sci. Technol. 48, 517–525.

O'Brien, J.W., Banks, A.P.W., Novic, A.J., Mueller, J.F., Jiang, G., Ort, C., et al., 2017. Impact of in-sewer degradation of pharmaceutical and personal care products (PPCPs) population markers on a population model. Environ. Sci. Technol. 51, 3815–3823.

ODPHP, 2017. History & Development of Healthy People. Office of Disease Prevention and Health Promotion, U.S. Department of Health and Human Services, Washington, DC Available on:. https://www.healthypeople.gov/2020/about/History-and-Development-of-Healthy-People, Accessed date: 13 July 2017.

Ort, C., Lawrence, M.G., Reungoat, J., Mueller, J.F., 2010. Sampling for PPCPs in wastewater systems: comparison of different sampling modes and optimization strategies. Environ. Sci. Technol. 44, 6289–6296.

Ort, C., Banta-Green, C., Been, F., Bijlsma, L., Catiglioni, S., Emke, E., et al., 2013. Sewage Analyses as an Early Detection System for Diseases and Indicator of Various Public Health Aspects with a Focus on Illicit Drugs. :p. 4 Davos, Switzerland. Available on:. http://ecite.utas.edu.au/86753, Accessed date: 13 July 2017.

Otterpohl, R., 2002. Options for alternative types of sewerage and treatment systems directed to improvement of the overall performance. Water Sci. Technol. 45, 149–158.

Pant, S., Hilton, H., Burczynski, M.E., 2012. The multifaceted exosome: biogenesis, role in normal and aberrant cellular function, and frontiers for pharmacological and biomarker opportunities. Biochem. Pharmacol. 83, 1484–1494.

Phung, D., Mueller, J., Lai, F.Y., O'Brien, J., Dang, N., Morawska, L., et al., 2017. Can wastewater-based epidemiology be used to evaluate the health impact of temperature? —an exploratory study in an Australian population. Environ. Res. 156, 113–119.

Piroddi, M., Albini, A., Fabiani, R., Giovannelli, L., Luceri, C., Natella, F., et al., 2016. Nutrigenomics of extra-virgin olive oil: a review. Biofactors 43, 17–41.

Poletti, A., Treville, A., 2016. Nano and micro-sensors: real time monitoring for the smart and sustainable city. Chem. Eng. Trans. 47, 1–6.

Prichard, J., Hall, W., de Voogt, P., Zuccato, E., 2014. Sewage epidemiology and illicit drug research: the development of ethical research guidelines. Sci. Total Environ. 472, 550–555.

Prichard, J., Hall, W., Zuccato, E., Voogt, P.D., Voulvoulis, N., Kummerer, K., et al., 2015. Ethical Research Guidelines for Wastewater-based Epidemiology and Related Fields, Version 1. Sewage Analysis CORe group Europe (SCORE):p. 13 Available on:. http://www.emcdda.europa.eu/attachements.cfm/att_251665_EN_WBE%20ethical%20guidelines%20FINAL%20March%202016%20.docx, Accessed date: 13 July 2017.

Ramin, P., Libonati Brock, A., Polesel, F., Causanilles, A., Emke, E., de Voogt, P., et al., 2016. Transformation and sorption of illicit drug biomarkers in sewer systems: understanding the role of suspended solids in raw wastewater. Environ. Sci. Technol. 50, 13397–13408.

Reddy, G.S., Tserng, K.Y., 1989. Calcitroic acid, end product of renal metabolism of 1,25-dihydroxyvitamin $D_3$ through C-24 oxidation pathway. Biochemistry 28, 1763–1769.

Reid, M.J., Langford, K.H., Mørland, J., Thomas, K.V., 2011. Analysis and interpretation of specific ethanol metabolites, ethyl sulfate, and ethyl glucuronide in sewage effluent for the quantitative measurement of regional alcohol consumption. Alcohol. Clin. Exp. Res. 35, 1–7.

Reis-Castro, L., 2017. The Underworlds Project and the "Collective Microbiome": Mining Biovalue From Sewage. In: Pavone, V., Goven, J. (Eds.), Bioeconomies: Life, Technology, and Capital in the 21st Century. Palgrave Macmillan, Cham, Switzerland, pp. 105–127.

Rico, M., Andrés-Costa, M.J., Picó, Y., 2017. Estimating population size in wastewater-based epidemiology. Valencia metropolitan area as a case study. J. Hazard. Mater. 323 (Part A), 156–165.

Ross, A.B., 2012. Present status and perspectives on the use of alkylresorcinols as biomarkers of wholegrain wheat and rye intake. J. Nutr. Metab., 462967 https://doi.org/10.1155/2012/462967:12.

Rousis, N.I., Gracia-Lor, E., Zuccato, E., Bade, R., Baz-Lomba, J.A., Castrignanò, E., et al., 2017. Wastewater-based epidemiology to assess pan-European pesticide exposure. Water Res. 121, 270–279.

Ryan, D., Robards, K., Prenzler, P.D., Kendall, M., 2011. Recent and potential developments in the analysis of urine: a review. Anal. Chim. Acta 684, 17–29.

Ryu, Y., Reid, M.J., Thomas, K.V., 2015. Liquid chromatography–high resolution mass spectrometry with immunoaffinity clean-up for the determination of the oxidative stress biomarker 8-iso-prostaglandin F2alpha in wastewater. J. Chromatogr. A 1409, 146–151.

Ryu, Y., Gracia-Lor, E., Bade, R., Baz-Lomba, J.A., Bramness, J.G., Castiglioni, S., et al., 2016. Increased levels of the oxidative stress biomarker 8-iso-prostaglandin $F_{2\alpha}$ in wastewater associated with tobacco use. Sci. Rep. 6, 39055.

Saah, A.J., Hoover, D.R., 1997. "Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings. Ann. Intern. Med. 126, 91–94.

Sagner, M., McNeil, A., Puska, P., Auffray, C., Price, N.D., Hood, L., et al., 2016. The P4 health spectrum—a predictive, preventive, personalized and participatory continuum for promoting healthspan. Prog. Cardiovasc. Dis. 59, 506–521.

Santos, J.M., Jurban, M., Kim, H., 2015. Could sewage epidemiology be a strategy to assess lifestyle and wellness of a large scale population? Med. Hypotheses 85, 408–411.

Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C., Dragsted, L.O., Draper, J., et al., 2014. The food metabolome: a window over dietary exposure. Am. J. Clin. Nutr. 99, 1286–1308.

Schepers, E., Speer, T., Bode-Böger, S.M., Fliser, D., Kielstein, J.T., 2014. Dimethylarginines ADMA and SDMA: the real water-soluble small toxins? Semin. Nephrol. 34, 97–105.

Schneider, M., Thoss, G., Hübner-Parajsz, C., Kientsch-Engel, R., Stahl, P., Pischetsrieder, M., 2004. Determination of glycated nucleobases in human urine by a new monoclonal antibody specific for $N^2$-carboxyethyl-2′-deoxyguanosine. Chem. Res. Toxicol. 17, 1385–1390.

Selleck, M.J., Senthil, M., Wall, N.R., 2017. Making meaningful clinical use of biomarkers. Biomark. Insights 12, 1–7.

Senta, I., Gracia-Lor, E., Borsotti, A., Zuccato, E., Castiglioni, S., 2015. Wastewater analysis to monitor use of caffeine and nicotine and evaluation of their metabolites as biomarkers for population size assessment. Water Res. 74, 23–33.

Servillo, L., Giovane, A., Cautela, D., Castaldo, D., Balestrieri, M.L., 2013. The methylarginines NMMA, ADMA, and SDMA are ubiquitous constituents of the main vegetables of human nutrition. Nitric Oxide 30, 43–48.

Simha, P., Ganesapillai, M., 2017. Ecological sanitation and nutrient recovery from human urine: how far have we come? A review. Sustain. Environ. Res. 27, 107–116.

Spahr, C.S., Davis, M.T., McGinley, M.D., Robinson, J.H., Bures, E.J., Beierle, J., et al., 2001. Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest. Proteomics 1, 93–107.

Straub, J.O., 2016. Reduction in the environmental exposure of pharmaceuticals through diagnostics, personalised healthcare and other approaches. A mini review and discussion paper. Sustain. Chem. Pharm. 3, 1–7.

Street, J.M., Koritzinsky, E.H., Glispie, D.M., Star, R.A., Yuen, P.S.T., 2017. Urine exosomes: an emerging trove of biomarkers. In: Gregory, S.M. (Ed.), Advances in Clinical Chemistry. 78. Elsevier, pp. 103–122 (chapter 3).

Thai, P.K., O'Brien, J., Jiang, G., Gernjak, W., Yuan, Z., Eaglesham, G., et al., 2014. Degradability of creatinine under sewer conditions affects its potential to be used as biomarker in sewage epidemiology. Water Res. 55, 272–279.

Thomaidis, N.S., Gago Ferrero, P., Ort, C., Maragou, N.C., Alygizakis, N.A., Borova, V.L., et al., 2016. Reflection of socio-economic changes in wastewater: licit and illicit drug use patterns. Environ. Sci. Technol. 50, 10065–10072.

Thomas, K.V., Amador, A., Baz-Lomba, J.A., Reid, M., 2017. Use of mobile device data to better estimate dynamic population size for wastewater-based epidemiology. Environ. Sci. Technol. 51, 11363–11370.

Toussaint, M.J., Hogarth, C.J., Nguyen, T.K.A., Gruys, E., 2004. Acute phase variables to assess health in various species. In: Madec, F., Clement, G. (Eds.), Animal Production in Europe: the Way Forward in a Changing World. Proceedings of the in-between Congress of the ISAH. 1. International Society for Animal Hygiène—Saint-Malo, St Malo, France, pp. 119–120.

van der Greef, J., van Wietmarschen, H., van Ommen, B., Verheij, E., 2013. Looking back into the future: 30 years of metabolomics at TNO. Mass Spectrom. Rev. 32, 399–415.

van't Erve, T.J., Kadiiska, M.B., London, S.J., Mason, R.P., 2017. Classifying oxidative stress by $F_2$-isoprostane levels across human diseases: a meta-analysis. Redox Biol. 12, 582–599.

Vermeulen, M., van den Berg, R., Freidig, A.P., van Bladeren, P.J., Vaes, W.H.J., 2006. Association between consumption of cruciferous vegetables and condiments and excretion in urine of isothiocyanate mercapturic acids. J. Agric. Food Chem. 54, 5350–5358.

Vlassov, A.V., Magdaleno, S., Setterquist, R., Conrad, R., 2012. Exosomes: current knowledge of their composition, biological functions, and diagnostic and therapeutic potentials. Biochim. Biophys. Acta 1820, 940–948.

Vogt, H., Hofmann, B., Getz, L., 2016. The new holism: P4 systems medicine and the medicalization of health and life itself. Med. Health Care Philos. 19, 307–323.

Waikar, S.S., Sabbisetti, V.S., Bonventre, J.V., 2010. Normalization of urinary biomarkers to creatinine during changes in glomerular filtration rate. Kidney Int. 78, 486–494.

Waris, S., Winklhofer-Roob, B.M., Roob, J.M., Fuchs, S., Sourij, H., Rabbani, N., et al., 2015. Increased DNA dicarbonyl glycation and oxidation markers in patients with type 2 diabetes and link to diabetic nephropathy. J. Diabetes Res. 2015, 915486.

Warrack, B.M., Hnatyshyn, S., Ott, K.-H., Reily, M.D., Sanders, M., Zhang, H., et al., 2009. Normalization strategies for metabonomic analysis of urine samples. J. Chromatogr. B 877, 547–552.

Weaver, V.M., Kotchmar, D.J., Fadrowski, J.J., Silbergeld, E.K., 2016. Challenges for environmental epidemiology research: are biomarker concentrations altered by kidney function or urine concentration adjustment? J. Expo. Sci. Environ. Epidemiol. 26, 1–8.

Westgate, P.J., 2009. Characterization of Proteins in Effluents from Three Wastewater Treatment Plants that Discharge to the Connecticut River. Masters dissertation. 42. University of Massachusetts Amherst, Environmental & Water Resources Engineering Masters Projects:p. 53 Available on:. http://scholarworks.umass.edu/cee_ewre/42, Accessed date: 13 July 2017.

Westgate, P.J., Park, C., 2010. Evaluation of proteins and organic nitrogen in wastewater treatment effluents. Environ. Sci. Technol. 44, 5352–5357.

Wild, C.P., 2005. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol. Biomark. Prev. 14, 1847–1850.

Wild, C.P., 2012. The exposome: from concept to utility. Int. J. Epidemiol. 41, 24–32.

Wilson, D.H., Rissin, D.M., Kan, C.W., Fournier, D.R., Piech, T., Campbell, T.G., et al., 2016. The Simoa HD-1 analyzer: a novel fully automated digital immunoassay analyzer with single-molecule sensitivity and multiplexing. J. Lab. Autom. 21, 533–547.

Winker, M., Tettenborn, F., Faika, D., Gulyas, H., Otterpohl, R., 2008. Comparison of analytical and theoretical pharmaceutical concentrations in human urine in Germany. Water Res. 42, 3633–3640.

Yang, Z., Anglès d'Auriac, M., Goggins, S., Kasprzyk-Hordern, B., Thomas, K.V., Frost, C.G., et al., 2015. A novel DNA biosensor using a ferrocenyl intercalator applied to the potential detection of human population biomarkers in wastewater. Environ. Sci. Technol. 49, 5609–5617.

Yoshida, M., Mikami, T., Higashi, K., Saiki, R., Mizoi, M., Fukuda, K., et al., 2012. Inverse correlation between stroke and urinary 3-hydroxypropyl mercapturic acid, an acrolein-glutathione metabolite. Clin. Chim. Acta 413, 753–759.

Yoshida, M., Higashi, K., Kuni, K., Mizoi, M., Saiki, R., Nakamura, M., et al., 2015. Distinguishing mild cognitive impairment from Alzheimer's disease with acrolein metabolites and creatinine in urine. Clin. Chim. Acta 441, 115–121.

Yu, O.B., Arnold, L.A., 2016. Calcitroic acid—a review. ACS Chem. Biol. 11, 2665–2672.

Zheng, Q.-D., Lin, J.-G., Pei, W., Guo, M.-X., Wang, Z., Wang, D.-G., 2017. Estimating nicotine consumption in eight cities using sewage epidemiology based on ammonia nitrogen equivalent population. Sci. Total Environ. 590–591, 226–232.

Zimmerman, D.R., Reinhardt, T.A., Kremer, R., Beitz, D.C., Reddy, G.S., Horst, R.L., 2001. Calcitroic acid is a major catabolic metabolite in the metabolism of $1\alpha$-dihydroxyvitamin $D_2$. Arch. Biochem. Biophys. 392, 14–22.

Zuccato, E., Chiabrando, C., Castiglioni, S., Calamari, D., Bagnati, R., Schiarea, S., et al., 2005. Cocaine in surface waters: a new evidence-based tool to monitor community drug abuse. Environ. Health J. 4.