Science of the Total Environment 619–620 (2018) 748–764

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

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

Christian G. Daughton

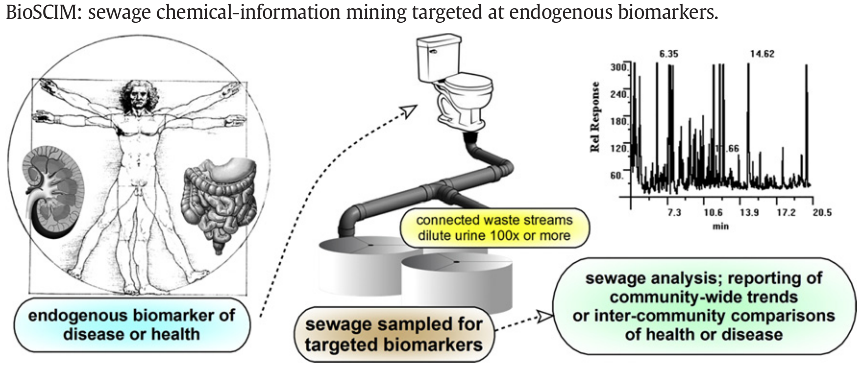
Environmental Futures Analysis Branch, Systems Exposure Division, National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, 944 East Harmon Avenue, Las Vegas 89119, NV, USA

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.



ARTICLE INFO

Article history:
Received 10 October 2017
Received in revised form 8 November 2017
Accepted 8 November 2017
Available online xxxx

Editor: D. Barcelo

Keywords:
Epidemiology
Public health
Endogenous biomarker
Urine
Exosomes
Exposome

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.

Abbreviations: API, acute-phase index; APPI, acute-phase protein index; APP, acute-phase proteins; APR, acute-phase response; BioSCIM, Sewage Chemical-Information Mining targeted at endogenous biomarkers; BoH, biomarkers of health (or wellness); d:hBR, disease:health biomarker ratio; P4H, predictive, preventive, personalized, and participatory healthcare; SCIM, Sewage Chemical-Information Mining; WWTPs, wastewater treatment plants.

E-mail address: daughton.christian@epa.gov.

https://doi.org/10.1016/j.scitotenv.2017.11.102

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 nonsteroidal 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