



HHS Public Access

Author manuscript

Metabolomics. Author manuscript; available in PMC 2015 April 23.

Published in final edited form as:

Metabolomics. 2014 April 1; 10(2): 176–178.

Navigating the road ahead: addressing challenges for use of metabolomics in epidemiology studies

Majda Haznadar,

Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA

Padma Maruvada,

Division of Digestive Diseases and Nutrition, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD 20817, USA

Eliza Mette,

Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD 20892, USA

John Milner,

Agricultural Research Service, United States Department of Agriculture, Beltsville, MD 20705, USA

Steven C. Moore,

Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20892, USA

Holly L. Nicastro,

Division of Cancer Prevention, National Cancer Institute, Rockville, MD 20892, USA

Joshua N. Sampson,

Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20892, USA

L. Joseph Su,

Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD 20892, USA

Mukesh Verma, and

Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD 20892, USA

Krista A. Zanetti

Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD 20892, USA

Abstract

© Springer Science+Business Media New York 2014

zanettik@mail.nih.gov.

Majda Haznadar, Padma Maruvada, Eliza Mette, John Milner, Steven C. Moore, Holly L. Nicastro, Joshua N. Sampson, L. Joseph Su, and Mukesh Verma have contributed equally to this work.

Metabolomics platforms allow for the measurement of hundreds to thousands of unique small chemical entities, as well as offer extensive coverage of metabolic markers related to obesity, diet, smoking, and other exposures of high interest to health scientists. Nevertheless, its potential use as a tool in population-based study design has not been fully explored. As the field of metabolomics continues to mature, and in part, accelerate through the National Institutes of Health (NIH) investment of 65 million in the Common Fund's Metabolomics Program (<https://commonfund.nih.gov/metabolomics/index>), it is time to consider those challenges most pertinent to epidemiologic studies.

1 Study-level challenges

To design and conduct high quality studies, at minimum the investigator must identify the general biological pathways of interest, the number of study participants and identify metabolomic endpoints a priori. Metabolomics provides a broad assessment of biology; therefore, investigators must determine in advance the general biological pathways and scientific questions of interest. This is because the biological pathways ascertained are mostly specific to the metabolomics platform and biospecimen used. For example, blood samples are excellent for assessing metabolism related to amino acids, fatty acids, and carbohydrates, but may be relatively weak, compared to urine, for assessing exposure to environmental endocrine disruptors, such as phthalate and bis-phenol-A, heavy metals, such as arsenic, or drug metabolism, such as alcohol and pain management medications.

Once the relevant biological pathways are identified, investigators must determine the appropriate study design. Here, we briefly review the pros and cons of two of the more common study designs, case-control and nested case-control, as they pertain to metabolomics. In a case-control study, samples are collected at the time of diagnosis; whereas in a nested case-control study, samples are collected prospectively, as part of a cohort study, prior to diagnosis and then followed until the clinical endpoint has been achieved. Case-control studies currently predominate in metabolomics research, possibly reflecting that samples from these studies are less costly and/or easier to obtain and provide distinct metabolic profiles between the treatment groups. In addition, because samples are collected at the time of disease onset in case-control studies, biomarkers of the disease itself may be present, which increases the likelihood of detecting unique markers that could be used for screening. Finally, metabolite-disease associations are likely to be stronger in case-control studies than in nested case-control studies, due to the proximity in time of sample collection to disease. Thus, for a fixed sample size, case-control studies may be better powered to detect associations. Overall, due to their lower expense and anticipated stronger effect sizes, case-control studies may be especially useful for exploratory analyses aimed at testing hypotheses of whether associations are evident for a given disease, and the number of potential associations.

Despite these advantages, case-control studies are much more likely to be affected by bias than nested case-control studies (Ernester 1994; Broadhurst and Kell 2006). Of particular concern is the potential for reverse causality. Typically, most investigators are interested in identifying etiologic factors that precede disease and increase the risk of the disease occurring but, in a case-control study, many of the metabolite-disease associations could be

the result of disease and may be of little intrinsic interest, e.g. statin metabolites may be elevated in people who have heart disease. Also, associations in a case–control study may occur due to study design artifacts. For example, if blood samples are drawn for cases in a fasted state during a clinical visit, and blood samples are drawn for controls in a non-fasted state during a home visit, then metabolite-disease associations may be identified, but many of them would simply reflect the difference in metabolite levels due to fasting status (Sampson et al. 2013). Case–control studies are also susceptible to selection bias, meaning that controls may not be representative of the source population that gives rise to the cases (Ernester 1994). Nevertheless, such investigations still often provide valuable insights for follow-up studies.

Perhaps the most difficult challenge is determining the appropriate number of study participants and obtaining the requisite sample size. In many cases, required sample sizes may be large. One reason is that in metabolomics it is common to examine hundreds of metabolites in relation to a disease outcome. To avoid false positives, correction for multiple testing must be done, such as a Bonferroni or false discovery rate adjustment. In theory, reducing the number of multiple tests by focusing on metabolites in just one biological pathway could help mitigate this loss in statistical power. However, such power comes at the high cost of omitting valuable data.

Additionally, effect sizes, e.g. odds ratios, may be weak, particularly if biospecimens were prospectively collected. In cancer epidemiology for example, there are only a few biomarkers for which the disease odds ratio are greater than 2.0 when comparing the top versus the bottom quartiles, including well-known examples such as the association between estrogen and breast cancer (Schairer et al. 2000) and aflatoxins and liver cancer (Qian et al. 1994). Such strong associations are the exception, not the rule (Berry 2012; Diamandis 2010). Thus, cancer epidemiology studies should be designed to detect small effect sizes. Additionally, integration of other “omic” technologies should be considered, as such technologies may help to better characterize the metabolic phenotype and to identify susceptible subpopulations, possibly resulting in more refined models with larger effect sizes.

Often the goal in epidemiology is to relate “usual” levels of a biomarker, such as blood pressure, vitamin D levels, or nicotine levels, with risk of disease, where usual is sometimes defined as the average concentration as a function of time, possibly the preceding month or perhaps a year. However, metabolites within an individual may substantially vary with time; therefore, metabolite levels from a single biospecimen may not reflect usual exposure or metabolic characteristics. This measurement error can cause associations to become attenuated, and requires a larger sample size to compensate (Sampson et al. 2013).

Finally, sample storage and handling require careful planning. Metabolite levels have been found to be extremely sensitive to numerous steps during sample allocation, far more so than other biomarkers. Therefore, in order to be comparable, samples must have identical histories in terms of storage and processing, even seeming benign procedures. This requires careful planning for samples that are collected as part of large epidemiological studies, as they are often used for multiple purposes.

2 Field-level challenges

While no single group can tackle the field-level challenges that researchers face, addressing standardized protocols, data storage and sharing, and availability of standard reference materials will help to advance the use of metabolomics in population-based studies. For clinical applications, molecular species need to be identified beyond any doubt; therefore, the availability of robust standards for identification is critical. Likewise, there is an indisputable need for standardization of metabolomic assays to be applied uniformly across laboratories.

In order to address the need for a database, encourage data sharing across the community and enhance abilities to compare future results between distinct studies, the NIH Common Fund's Metabolomics Program recently awarded funds to establish the Metabolomics Data Repository and Coordination Center, which is developing protocols for minimum metadata requirements for posting data into the repository. It has also recently launched a portal for metabolomics resources (<http://www.metabolomicsworkbench.org/>). Additionally, there are efforts of this nature taking place in Europe, including the Coordination of Standards in Metabolomics (COSMOS) (<http://www.cosmos-fp7.eu/>) and The European Bioinformatics Institute's (EBI) MetaboLights (<http://www.ebi.ac.uk/metabolights/>).

Additionally, the NIH has recently created a pooled plasma reference set in collaboration with the National Institute of Standards and Technology (NIST) (<http://srm1950.nist.gov/>). While the pooled specimens may provide information on the variety of metabolites in a given specimen, they are of limited value in providing information on individual variation in levels of a metabolite and the range of metabolites that may be clinically relevant. Metabolomics reference standards identified by the research community will be synthesized by the NIH Common Fund's contract mechanism as well. However, there remains a need for the collection of well-annotated and characterized diverse biological specimens, including urine and serum, to determine the normal metabolite profile, ranges of metabolites, and modulation of metabolites in pathophysiological conditions.

3 Future directions

The full potential for the advancement of metabolomics to epidemiology studies will require overcoming both study-level and field-level challenges. Study-level challenges will need to be addressed by individual investigators and require a clear understanding of the principles of epidemiological study design. Collaborative multidisciplinary teams with biochemists, epidemiologists, geneticists and biostatisticians will be needed to assure that studies are designed properly, the correct platforms are used, and the data is analyzed appropriately.

Field-level challenges must be addressed by the scientific community with dedicated funding to tackle the lack of standardized protocols, data storage and sharing, and availability of standard reference materials in the field. Some efforts related to data storage and sharing issues are addressed through the Metabolomics Data Repository and Coordination Center and EBI's MetaboLights (Salek et al. 2013); however, there have been limited efforts for addressing the need for standardized protocols or standard reference materials. If these field-level issues can be addressed, it will allow increased and more

effective use of metabolomic profiling in population-based studies for biomedical and public health research.

Overall, there are exciting opportunities for using metabolomics in epidemiological investigations, despite the significant challenges.

References

- Berry D. Multiplicities in cancer research: Ubiquitous and necessary evils. *Journal of the National Cancer Institute*. 2012; 104:1124–1132. [PubMed: 22859849]
- Broadhurst DI, Kell DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics*. 2006; 2:171–196.
- Diamandis EP. Cancer biomarkers: Can we turn recent failures into success? *Journal of the National Cancer Institute*. 2010; 102:1462–1467. [PubMed: 20705936]
- Ernster VL. Nested case-control studies. *Preventive Medicine*. 1994; 23:587–590. [PubMed: 7845919]
- Qian GS, Ross RK, Yu MC, Yuan JM, Gao YT, Henderson BE, et al. A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China. *Cancer Epidemiology, Biomarkers and Prevention*. 1994; 3:3–10.
- Salek RM, Steinbeck C, Goodacre R, Viant MR, Dunn WB. The role of reporting standards for metabolite annotation and identification in metabolomic studies. *GigaScience*. 2013; 2:13. [PubMed: 24131531]
- Sampson JN, Boca SM, Shu XO, Stolzenberg-Soloman RZ, Matthews CE, Hsing AW, et al. Metabolomics in epidemiology: Sources of variability in metabolite measurements and implications. *Cancer Epidemiology, Biomarkers and Prevention*. 2013; 22:631–640.
- Schairer C, Lubin J, Troisi R, Sturgeon S, Brinton L, Hoover R. Menopausal estrogen and estrogen-progestin replacement therapy and breast cancer risk. *Journal of the American Medical Association*. 2000; 283:485–491. [PubMed: 10659874]