



Published in final edited form as:

Metabolomics. ; 15(1): 9. doi:10.1007/s11306-018-1468-z.

Metabolomics in epidemiologic research: challenges and opportunities for early-career epidemiologists

Eline H. van Roekel¹, Erikka Lofffield², Rachel S. Kelly³, Oana A. Zeleznik³, Krista A. Zanetti⁴

¹Department of Epidemiology, GROW School for Oncology and Developmental Biology, Maastricht University, P.O. Box 616, 6200 MD Maastricht, Netherlands ²Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, USA ³Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, USA ⁴Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, USA

Abstract

Background—The application of metabolomics to epidemiologic studies is increasing.

Aim of Review—Here, we describe the challenges and opportunities facing early-career epidemiologists aiming to apply metabolomics to their research.

Key Scientific Concepts of Review—Many challenges inherent to metabolomics may provide early-career epidemiologists with the opportunity to play a pivotal role in answering critical methodological questions and moving the field forward. Although generating large-scale high-quality metabolomics data can be challenging, data can be accessed through public databases, collaboration with senior researchers or participation within interest groups. Such efforts may also assist with obtaining funding, provide knowledge on training resources, and help early-career epidemiologists to publish in the field of metabolomics.

Keywords

Metabolomics; Epidemiology; Early-career scientists; Challenges; Opportunities

1 Background: metabolomics in the epidemiologic context

Recent advances in metabolomic technologies, analytical methods and data processing tools have made metabolomic profiling of large-scale population-based studies possible (Bictash

Eline H. van Roekel, eline.vanroekel@maastrichtuniversity.nl.

Author contributions

All authors wrote the manuscript. All authors read and approved the manuscript.

Conflict of interest All authors declare that they do not have conflict of interest.

Research involving human and/or animal participants This article does not contain any studies with human and/or animal participants performed by any of the authors.

et al. 2010). Consequently, the application of metabolomic technologies to epidemiologic studies is a growing research area with the potential to better characterize exposures, detect early markers of disease, improve diagnosis of disease, track response to treatment or disease progression, and understand disease etiology (Liesenfeld et al. 2013, Su et al. 2014; Tzoulaki et al. 2014). Over the last decade, the number of metabolomics-focused epidemiologic studies has steadily increased (Fig. 1), highlighting the growing scientific interest in this research area.

Metabolomics provides a biological and mechanistic facet to research that is often not achievable with more classical epidemiologic studies which rely on questionnaire or environmental exposure data. The potential of the metabolome to reflect genetic background, environmental factors and phenotypes make it a complex but powerful tool for discovery. The experimental study of metabolism can be separated into two analytical chemical strategies, targeted and untargeted approaches (Liesenfeld et al. 2013; Tzoulaki et al. 2014). Targeted approaches focus on a small number of pre-defined metabolites in a hypothesis-based approach; these approaches have been referred to as targeted metabolomics and clinical chemistry. Untargeted approaches provide a global investigation of metabolites in a biological system in a hypothesis-free manner; these approaches have been referred to as metabolomics or metabonomics. Untargeted approaches, provide more potential for discovery of novel metabolic associations and disease pathways; however, untargeted analyses are more costly and in particular *post hoc* metabolite identification may be difficult (Tzoulaki et al. 2014). As a result, most epidemiological studies to date have used targeted analysis. Nevertheless, the application of untargeted methods in large-scale studies has seen an increase in recent years.

Metabolomics has been applied in epidemiological studies with a variety of study designs. To date, a majority of metabolomic-based epidemiologic studies have been cross-sectional or case-control studies allowing for the comparison of metabolomic phenotypes of individuals by disease or exposure status to gain insight into potential metabolic differences. Prospective studies generally require large sample sizes and extended follow-up time, however they provide a powerful opportunity for the identification of biomarkers of susceptibility or pre-clinical disease and for gaining insight into disease etiology (Zanetti et al. 2014). Consequently, an increasing number of prospective studies with efficient nested-designs, namely nested case-control and case-cohort studies, are now being conducted. Table 1 provides an overview of epidemiological study designs and definitions, along with examples of epidemiological studies using metabolomics data. Below, we describe some of the applications of metabolomics in different fields of epidemiological research.

Many epidemiological studies have used metabolomic data to identify biomarkers of chronic metabolic diseases such as type II diabetes (Wang et al. 2011; Wang-Sattler et al. 2012; Floegel et al. 2013; Menni et al. 2013), cardiovascular disease (Wurtz et al. 2015), hypertension (Holmes et al. 2008; Zheng et al. 2013b), and heart failure (Zheng et al. 2013a). In cancer research, studies have also mostly focused on identifying potential noninvasive biomarkers of disease risk and prognosis (Nishiumi et al. 2012; Liesenfeld et al. 2013; Cross et al. 2014; Mathe et al. 2014; Fages et al. 2015; Ke et al. 2015; Mondul et al. 2015; Huang et al. 2016; Stepien et al. 2016). Metabolomics has also been applied in

epidemiologic studies of non-chronic diseases aimed at identifying diagnostic and prognostic biomarkers, e.g. infectious diseases such as Lyme disease (Molins et al. 2015) and dengue fever (Cui et al. 2016). In addition, another major area of interest has been metabolomic profiling of disease risk factors such as adiposity (Moore et al. 2014; Ho et al. 2016; Carayol et al. 2017) and diet (Scalbert et al. 2009, Su et al. 2014). Recent studies have extended this research by using prospective study data to identify relationships between exposure-related metabolites and cancer outcomes (Guertin et al. 2015; Playdon et al. 2017; Moore et al. 2018). Novel statistical approaches for mediation analysis (Assi et al. 2015) may further our understanding of the intermediate role of metabolites in exposure-disease associations.

2 Challenges and opportunities for early-career epidemiologists

The increasing application of metabolomics to epidemiologic studies, as well as the speed of technical and statistical developments, creates unique challenges and opportunities for early-career epidemiologists (here defined as up to 5 years post-PhD), who are trying to enter this research field. Key challenges include the technical and analytical issues inherent to all metabolomics research, which can be even more critical in large-scale epidemiologic studies. Further, early-career epidemiologists in metabolomics may also face specific practical and logistic concerns, which can be more pronounced when experience with metabolomics-based research is still limited. In this article, we will describe the challenges and opportunities facing early-career epidemiologists in the metabolomics field. The first four authors of this manuscript (EvR, EL, RK, OZ) are early-career researchers according to the previously mentioned definition; this manuscript has been written based on their personal experience, complemented with the perspective of a senior scientist with a long-standing experience in this field (KZ).

2.1 Technical and statistical challenges

To be able to generate and use metabolomics data in an epidemiologic study, understanding the technical and statistical challenges is critical. Technical challenges typically arise during the sample collection, processing, and laboratory analysis phases, while the statistical challenges involve making statistical and, more crucially, biological sense of multidimensional, highly correlated and noisy datasets. The type of specimen, collection method, storage conditions, and profiling-related variables such as analytical batch can impact metabolomic profiles and add noise to the data (Tzoulaki et al. 2014; Cheng et al. 2017; Hernandez et al. 2017; Kirwan et al. 2018). Accounting for this can be particularly complex for epidemiological metabolomics studies as these are often based on samples which were not collected with a metabolomics analysis in mind; rather, researchers often use pre-existing samples from large-scale population-based studies or human biobanks. There is growing consensus in the field that standards to deal with technical challenges are vital, and a number of international collaborations are now working towards developing and implementing such standards (Dudzik et al. 2018). In particular, they have called for more for more accurate reporting protocols of all the steps taken before during and after data acquisition, to assess their influence. For example, the Metabolomics Standards Initiative and Chemical Analysis Working Group, are both working toward the development of

minimum reporting standards which they hope will become mandatory for publication in high ranking peer-reviewed journals (Sumner et al. 2007). Furthermore, an active voice within the metabolomics community has argued for the critical importance of including standard quality control samples in every analytical run in order to facilitate inter-study and inter-laboratory comparisons (Broadhurst et al. 2018). The need for standardization is particularly applicable to untargeted analysis, where methods have been less well-developed and *post hoc* metabolite identification may be difficult, as methods were not optimized for measuring specific metabolites (Tzoulaki et al. 2014).

There is a similar movement towards the development of standards for the statistical analysis of metabolomic data, particularly with relation to two of the most pressing issues, namely multiple testing (Saccenti et al. 2014) and missing data (Hrydziuszko 2012). These issues are especially relevant when working with metabolomic data, because of the large number of correlated metabolite variables, many of which are detectable in some but not all samples. Efforts are ongoing but to date, these have not moved beyond recommendations and no gold-standard approaches have been agreed upon. It is also unclear how statistical methods can best integrate the biological knowledge underpinning the pathway- and network-based structure of metabolites. Further, there are no standards for reporting results or metabolite nomenclature and when alternative metabolite names are reported in publications, it can be difficult to compare results across studies. Finally, the traditional statistical analysis approaches in epidemiology and metabolomics differ substantially. While epidemiologists mostly apply regression-based approaches to identify metabolites of interest, the metabolomics community largely relies on high dimensional data analysis methods such as principal component analysis, clustering and pathway analysis (Ren et al. 2015).

Technical and statistical challenges, though complex and multiple, are largely surmountable and provide unique opportunities for early-career epidemiologists to contribute to developing methodologies. Small but well-designed methods studies can answer a multitude of questions ranging from the impact of collection and storage methods (Loftfield et al. 2016) to biological variability of metabolites over time or across different sample types and platforms (Floegel et al. 2013; Townsend et al. 2013; Xiao et al. 2014; Carayol et al. 2017). These studies can typically be executed with limited resources and within a short period of time, making them feasible for early-career researchers who are interested in generating their own data. Such studies also afford early-career investigators unique insight into research methodology and study design that they can leverage when proposing future studies. Early-career investigators should also consider whether they can answer important questions with available data. For example, existing data may be used to develop novel statistical methods or visualization tools for improving data analysis, integration and interpretation (Fages et al. 2014; Alkawaa et al. 2017).

2.2 Obtaining metabolomics data

Generating metabolomics data, particularly untargeted data, in a large, epidemiologic study is generally costly. If a researcher has access to biospecimens from a prospective cohort study, a well-powered nested case-control study can be a less costly and effective study

design; however, funding or an academic collaboration with a metabolomics laboratory would still be required.

Gaining access to existing data is an efficient and cost effective way to gain experience in the field. The simplest option is represented by online databases, such as MetaboLights (<https://www.ebi.ac.uk/metabolights/>; mainly supported by the European Molecular Biology Laboratory, UK Medical Research Council and Wellcome Trust), and the National Institute of Health (NIH)-supported Metabolomics Workbench (<http://www.metabolomicsworkbench.org/>). Another way to gain access to data is to collaborate with a senior metabolomics researcher. There are generally more hypotheses than scientists to test them, and senior investigators may be particularly willing to host early-career epidemiologists with unique expertise, such as epidemiologic study design or statistical data analysis.

Joining a scientific society or special interest group represents another way to potentially gain access to available data and connect researchers with a shared interest in metabolomics from different fields, such as epidemiology, chemistry and bioinformatics. The Metabolomics Society (<http://metabolomicsociety.org/>) and COnsortium of METabolomics Studies (COMETS) (<https://epi.grants.cancer.gov/comets/>) are two such groups. The Metabolomics Society was founded in 2004 through joint efforts of several scientists representing various research areas, including metabolomics, metabolic flux analysis, biochemical modeling, and related informatics fields; membership is open to anyone. The Early-career Members Network (EMN) is a forum for early-career members within the Metabolomics Society that organizes regular workshops and webinars and provides an online social platform. COMETS is a prospective cohort study-centric partnership founded in 2014 with the aim to support and promote the study of metabolomics profiles associated with chronic diseases and health-related lifestyle factors such as adiposity. COMETS membership is open to scientists of participating studies, as well as for researchers working with metabolomics data in prospective cohort studies and early-career scientists within their teams (for membership possibilities, see: <https://epi.grants.cancer.gov/comets/>). COMETS recognizes the challenges faced by early-career investigators in this field and initiated an early-career scientists working group where early-career scientists can network, share experiences and problems in an organized setting and propose and develop small metabolomics-related projects.

2.3 Funding

Financial support for generating metabolomics data in a population-based study is mostly available via funding mechanisms open to senior scientists. Announcements such as that from the Next Generation Research Initiative of the NIH signal a shift towards research funding opportunities for early- and mid-career investigators. Early-career scientists may also be eligible for pilot study funding grants, such as those from the World Cancer Research Fund (<http://www.wcrf.org/int/research-we-fund/grant-programmes>), or for personal research grants for early-career researchers such as NIH Pathway to Independence Award (K99/R00) program (<https://grants.nih.gov/grants/guide/pa-files/PA-18-398.html>) or a European Marie Skłodowska-Curie Fellowship (<https://ec.europa.eu/research/>

[mariecurieactions/](#)). These grants can be an important first step in obtaining a grant and publication record, collaborating with a metabolomics laboratory, analyzing metabolomics data, and increasing one's visibility in the research field.

2.4 Other resources

Beyond funding, other resources are required to conduct a successful study and develop a career within a new field, including special interest groups and committees, and online resources for planning, conducting, analyzing and interpreting data. A key challenge for early-career epidemiologists in metabolomics may be the limited pool of experts and senior epidemiologists to act as mentors. In addition, the multi-disciplinary nature of metabolomics requires a team of experienced researchers with knowledge and skills from a variety of research fields, including biostatistics, bioinformatics, epidemiology, chemistry and biochemistry. Therefore early-career epidemiologists should be prepared to foster external collaborations as well as a broader and potentially global research network. Attending conferences and training workshops represents a rewarding and efficient way of identifying all available resources and of networking with senior scientists and other metabolomics researchers. One of the largest metabolomics conferences is the annual conference of the Metabolomics Society (<http://metabolomicssociety.org>). The Society's EMN offers a travel bursary program for early-career scientists for attending national or international metabolomics meetings.

Online and offline metabolomics training is offered, amongst others, by the NIH Common Fund Metabolomics Program Regional Comprehensive Metabolomics Resource Cores (<https://commonfund.nih.gov/metabolomics/training>) with one example being the University of Alabama at Birmingham (UAB) Metabolomics Workshop (<https://www.uab.edu/proteomics/metabolomics/workshop.php>), and by the European Bioinformatics Institute (<https://www.ebi.ac.uk/training>) and the Birmingham Metabolomics Training Centre of the University of Birmingham, UK (<https://www.birmingham.ac.uk/facilities/metabolomics-trainingcentre/index.aspx>). Recently, the European Metabolomics Training Coordination Group (<http://www.emtrag.eu>) was established to harmonize metabolomics training in Europe. These trainings are ideally suited to the needs of early-career epidemiologists wishing to leverage metabolomics in their research.

2.5 Publications

Publishing findings from a metabolomics study in the epidemiologic context can be hampered by the lack of standardized guidelines on data generation, analysis and interpretation. Publication of initial findings may be harder for early-career epidemiologists who are yet to develop a reputation among the metabolomics community. Writing a literature review may be a valuable means of gaining knowledge in one's area of interest as well as a publication record. Moreover, it can help to identify relevant research gaps and there are many areas in the field of metabolomics-based epidemiology for which robust systematic reviews are lacking. In addition, collaborating with other metabolomics researchers may be an important means to gain experience in the field and to increase the number of co-authored publications.

3 Conclusion

The application of metabolomics in epidemiologic studies is a growing research field. Its promise to increase our understanding of human health has been recognized by funding and research agencies such as the NIH and the International Agency for Research on Cancer (IARC). In comparison to other -omics fields, such as genomics, metabolomics research methods are less developed, which presents challenges, standardization being one of them. However, there are also opportunities for early-career epidemiologists to establish a career in the metabolomics field and potentially make seminal contributions to its advancement in population sciences. In fact, the multidisciplinary team science approach of metabolomics provides innumerable opportunities for early-career investigators trained in epidemiology to build collaborations and contribute to science in productive and innovative ways.

Funding

E.H. van Roekel was financially supported by Wereld Kanker Onderzoek Fonds (WKOF), as part of the World Cancer Research Fund International grant programme (Grant No. 2016/1620) and the GROW School for Oncology and Developmental Biology. E. Lofffield was supported by the Intramural Research Program of the National Institutes of Health, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Department of Health and Human Services. R.S. Kelly was supported by a Discovery Award from The US Department of Defense (Grant No. W81XWH-17-1-0533), and a grant from the US NIH (Grant No. 1R01HL123915-01). O.A. Zeleznik was supported by grants from the NIH (Grant Nos. CA087969, CA050385). K.A. Zanetti was supported by the Extramural Research Program of the National Institutes of Health, Division of Cancer Control and Populations Sciences, National Cancer Institute, Department of Health and Human Services.

References

- Alkawaa FM, et al. (2017). Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. *Journal of Proteome Research*, 17(1), 337–347. [PubMed: 29110491]
- Assi N, et al. (2015). A statistical framework to model the meeting-in-the-middle principle using metabolomic data: Application to hepatocellular carcinoma in the EPIC study. *Mutagenesis*, 30(6), 743–753. [PubMed: 26130468]
- Baldassarre ME, et al. (2018). Effectiveness and safety of a probiotic-mixture for the treatment of infantile colic: A double-blind, randomized, placebo-controlled clinical trial with fecal real-time PCR and nmr-based metabolomics analysis. *Nutrients*, 10(2), 195.
- Bictash M, et al. (2010). Opening up the “Black Box”: Metabolic phenotyping and metabolome-wide association studies in epidemiology. *Journal of Clinical Epidemiology*, 63(9), 970–979. [PubMed: 20056386]
- Broadhurst D, et al. (2018). Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics*, 14(6), 72. [PubMed: 29805336]
- Carayol M, et al. (2017). Blood metabolic signatures of body mass index: A targeted metabolomics study in the EPIC cohort. *Journal of Proteome Research*, 16(9), 3137–3146. [PubMed: 28758405]
- Cheng S, et al. (2017). Potential impact and study considerations of metabolomics in cardiovascular health and disease: A scientific statement from the American Heart Association. *Circulation Cardiovascular Genetics*, 10(2), e000032.
- Cross AJ, et al. (2014). A prospective study of serum metabolites and colorectal cancer risk. *Cancer*, 120(19), 3049–3057. [PubMed: 24894841]
- Cui L, et al. (2016). Serum metabolomics reveals serotonin as a predictor of severe dengue in the early phase of dengue fever. *PLoS Neglected Tropical Diseases*, 10(4), e0004607. [PubMed: 27055163]
- Derkach A, et al. (2017). Effects of dietary sodium on metabolites: The dietary approaches to stop hypertension (DASH)-sodium feeding study. *The American Journal of Clinical Nutrition*, 106(4), 1131–1141. [PubMed: 28855223]

- Dudzik D, et al. (2018). Quality assurance procedures for mass spectrometry untargeted metabolomics. A review. *Journal of Pharmaceutical and Biomedical Analysis*, 147, 149–173. [PubMed: 28823764]
- Fages A, et al. (2014). Investigating sources of variability in metabolomic data in the EPIC study: The principal component partial R-square (PC-PR2) method. *Metabolomics*, 10(6), 1074–1083.
- Fages A, et al. (2015). Metabolomic profiles of hepatocellular carcinoma in a European prospective cohort. *BMC Medicine*, 13, 242. [PubMed: 26399231]
- Floegel A, et al. (2013). Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes*, 62(2), 639–648. [PubMed: 23043162]
- Floegel A, et al. (2018). Serum metabolites and risk of myocardial infarction and ischemic stroke: A targeted metabolomic approach in two German prospective cohorts. *European Journal of Epidemiology*, 33(1), 55–66. [PubMed: 29181692]
- Guertin KA, et al. (2015). Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. *American Journal of Clinical Nutrition*, 101(5), 1000–1011.
- Hernandes VV, et al. (2017). A review of blood sample handling and pre-processing for metabolomics studies. *Electrophoresis*, 38(18), 2232–2241. [PubMed: 28543881]
- Ho JE, et al. (2016). Metabolomic profiles of body mass index in the framingham heart study reveal distinct cardiometabolic phenotypes. *PLoS ONE*, 11(2), e0148361. [PubMed: 26863521]
- Holmes E, et al. (2008). Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature*, 453(7193), 396–400. [PubMed: 18425110]
- Hrydziuszko O, & Viant MR (2012). Missing values in mass spectrometry based metabolomics: An undervalued step in the data processing pipeline. *Metabolomics*, 8(1), 161–174.
- Huang J, et al. (2016). Serum metabolomic profiling of prostate cancer risk in the prostate, lung, colorectal, and ovarian cancer screening trial. *British Journal of Cancer*, 115(9), 1087–1095. [PubMed: 27673363]
- Huang S, et al. (2016). Novel personalized pathway-based metabolomics models reveal key metabolic pathways for breast cancer diagnosis. *Genome Medicine*, 8(1), 34. [PubMed: 27036109]
- Ke C, et al. (2015). Large-scale profiling of metabolic dysregulation in ovarian cancer. *International Journal of Cancer*, 136(3), 516–526. [PubMed: 24895217]
- Kirwan JA, et al. (2018). Preanalytical processing and biobanking procedures of biological samples for metabolomics research: A white paper, community perspective (for “precision Medicine and pharmacometabolomics task group”-the metabolomics society initiative). *Clinical Chemistry*, 64(8), 1158–1182. [PubMed: 29921725]
- Liesenfeld DB, et al. (2013). Review of mass spectrometry-based metabolomics in cancer research. *Cancer Epidemiology Prevention Biomarkers*, 22(12), 2182–2201.
- Loftfield E, et al. (2016). Comparison of collection methods for fecal samples for discovery metabolomics in epidemiologic studies. *Cancer Epidemiology Prevention Biomarkers*, 25(11), 1483–1490.
- Mathe EA, et al. (2014). Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. *Cancer Research*, 74(12), 3259–3270. [PubMed: 24736543]
- Menni C, et al. (2013). Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes*, 62(12), 4270–4276. [PubMed: 23884885]
- Molins CR, et al. (2015). Development of a metabolic biosignature for detection of early Lyme disease. *Clinical Infectious Diseases*, 60(12), 1767–1775. [PubMed: 25761869]
- Mondul AM, et al. (2015). Metabolomic analysis of prostate cancer risk in a prospective cohort: The alpha-tocopherol, beta-carotene cancer prevention (ATBC) study. *International Journal of Cancer*, 137(9), 2124–2132. [PubMed: 25904191]
- Moore SC, et al. (2014). Human metabolic correlates of body mass index. *Metabolomics*, 10(2), 259–269. [PubMed: 25254000]
- Moore SC, et al. (2018). A metabolomics analysis of body mass index and postmenopausal breast cancer risk. *Journal of National Cancer Institute*, 110(6), 588–597.

- Nishiumi S, et al. (2012). A novel serum metabolomics-based diagnostic approach for colorectal cancer. *PLoS ONE*, 7(7), e40459. [PubMed: 22792336]
- Playdon MC, et al. (2017). Nutritional metabolomics and breast cancer risk in a prospective study. *The American Journal of Clinical Nutrition*, 106(2), 637–649. [PubMed: 28659298]
- Ren S, et al. (2015). Computational and statistical analysis of metabolomics data. *Metabolomics*, 11(6), 1492–1513.
- Saccenti E, et al. (2014). Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics*, 10(3), 361–374.
- Scalbert A, et al. (2009). Mass-spectrometry-based metabolomics: Limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics*, 5(4), 435–458. [PubMed: 20046865]
- Stepien M, et al. (2016). Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: Findings from a prospective cohort study. *International Journal of Cancer*, 138(2), 348–360. [PubMed: 26238458]
- Su LJ, et al. (2014). The use of metabolomics in population-based research. *Advances in Nutrition*, 5(6), 785–788. [PubMed: 25398741]
- Sumner LW, et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics*, 3(3), 211–221. [PubMed: 24039616]
- Townsend MK, et al. (2013). Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clinical Chemistry*, 59(11), 1657–1667. [PubMed: 23897902]
- Tzoulaki I, et al. (2014). Design and analysis of metabolomics studies in epidemiologic research: A primer on omic technologies. *American Journal of Epidemiology*, 180(2), 129–139. [PubMed: 24966222]
- Wang TJ, et al. (2011). Metabolite profiles and the risk of developing diabetes. *Nature Medicine*, 17(4), 448–453.
- Wang-Sattler R, et al. (2012). Novel biomarkers for pre-diabetes identified by metabolomics. *Molecular Systems Biology*, 8, 615. [PubMed: 23010998]
- Welsh P, et al. (2018). Circulating amino acids and the risk of macrovascular, microvascular and mortality outcomes in individuals with type 2 diabetes: Results from the advance trial. *Diabetologia*, 61(7), 1581–1591. [PubMed: 29728717]
- Wurtz P, et al. (2015). Metabolite profiling and cardiovascular event risk: A prospective study of 3 population-based cohorts. *Circulation*, 131(9), 774–785. [PubMed: 25573147]
- Xiao Q, et al. (2014). Sources of variability in metabolite measurements from urinary samples. *PLoS ONE*, 9(5), e95749. [PubMed: 24788433]
- Zanetti K, et al. (2014). The future of metabolomic profiling in population-based research: Opportunities and challenges. *Journal of Analytical and Bioanalytical Techniques*, 5, 203.
- Zheng Y, et al. (2013a). Associations between metabolomic compounds and incident heart failure among African Americans: The ARIC Study. *American Journal of Epidemiology*, 178(4), 534–542. [PubMed: 23788672]
- Zheng Y, et al. (2013b). Metabolomics and incident hypertension among blacks: The atherosclerosis risk in communities study. *Hypertension*, 62(2), 398–403. [PubMed: 23774226]

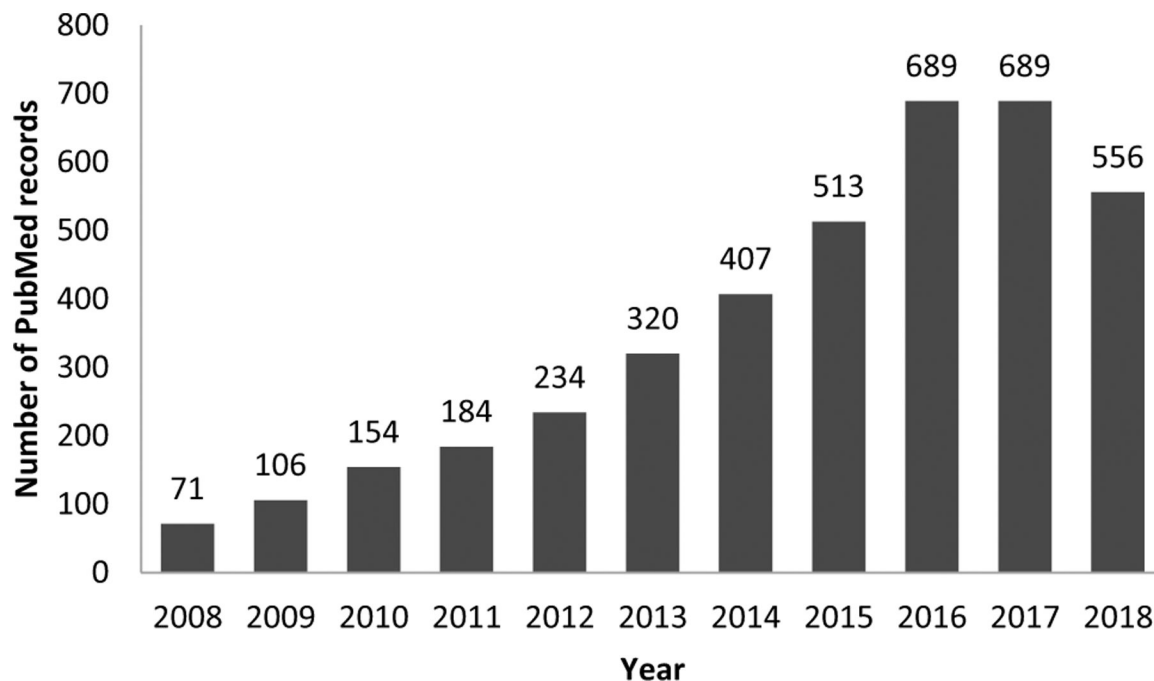


Fig. 1.

Yearly number of PubMed records including both keywords on epidemiology/population-based research and metabolomics in the period 2008–2018 (2018 number based on results up to October 6, 2018). Footnote: search performed as (“epidemiology” or “epidemiologic” or “population-based” or “observational” or “case-control” or “cohort” or “cross-sectional”) AND (“metabolome” or “metabonome” or “metabolomic” or “metabolomics” or “metabonomic” or “metabonomics” or “metabolic profile” or “metabolite profile” or “metabolic signature” or “glycomic” or “glycomics” or “lipidomic” or “lipidomics”); date of search: October 6, 2018

Table 1 Epidemiological study designs and definitions¹, and examples of epidemiological studies using metabolomics data by study design

Study design	Definition	Examples of metabolomics studies
Experimental Randomized controlled trial (RCT)	A trial, or epidemiologic experiment, on humans in which the investigator controls who is exposed and who is not exposed (as well as the dose, frequency, and duration of exposure) through random allocation of study participants to exposed and non-exposed (control) groups by chance (i.e., randomization)	(Derkach et al. 2017) (Baldassarre et al. 2018)
Observational Prospective cohort study	A study in which a specified group of disease-free participants, some of whom are exposed to the factor(s) of interest and some of whom are not exposed to the same factor(s) (i.e., comparison group), are followed for incidence of disease or death over a designated period of time	(Zheng et al. 2013 b) (Wurtz et al. 2015)
Case-control study	A study in which cases, those who have developed a given disease, and controls, those who have not developed the disease, are compared with respect to their history of past exposures of interest. The controls should be representative, with respect to the exposure(s) of interest, of the source population that gave rise to the cases	(Mathe et al. 2014) (Molins et al. 2015)
Case-cohort study	A nested study in which the source population consists of cohort participants who were disease-free at baseline. Controls are randomly selected from the source population at the start of follow-up (i.e., sub-cohort). Cases are new cases of the disease of interest that developed in the source population over the designated period of follow-up	(Welsh et al. 2018) (Floegel et al. 2018)
Nested case-control study	A case-control study nested within a cohort study in which the source population consists of cohort participants who were disease-free at baseline. Controls are randomly selected from those still at risk in the source population at the time a case occurs using density-based sampling. Cases are new cases of the disease of interest that developed over a designated period of follow-up	(Stepien et al. 2016) (Moore et al. 2018)
Cross-sectional study	A study of the distribution of an exposure and/or a disease in a well-defined population at a single point in time, giving a “snapshot” of the population	(Holmes et al. 2008) (Carayol et al. 2017)

¹Definitions are adapted from *Modern Epidemiology, Third Edition* by Rothman et al.