Studies in History and Philosophy of Biol & Biomed Sci 82 (2020) 101248

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biol & Biomed Sci

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

Making evidential claims in epidemiology: Three strategies for the study of the exposome

Stefano Canali

Institute for Philosophy, Leibniz Universität Hannover, Lange Laube 32, 30159, Hannover, Germany

ARTICLE INFO

Keywords:
Research practice
Data
Evidence
Exposome
Philosophy of epidemiology
Big data

ABSTRACT

How is scientific data used to represent phenomena and as evidence for claims about phenomena? In this paper, I propose that a specific type of claims – evidential claims – is involved in data practices to define and restrict the representational and evidential content of a dataset. I present an account of data practices in the epidemiology of the exposome based on the notion of evidential claims, which helps unpack the approaches, assumptions and warrants that connect different stages of research. I identify three different strategies to generate different types of evidential claims in this case. The macro strategy, which individuates the dataset that serves as the initial evidential space for research. The micro strategy, which is used to generate evidential claims about the microscopic and individual component of target phenomena. The association strategy, that uses evidence from the other strategies to identify a dataset as representation of the different levels and relations of exposure and disease. Differentiating between these strategies sheds light on the multi-faceted landscape of biomedical research on environment and health; and the roles of data and evidence in the process of inquiry.

1. Introduction

The health sciences have a long tradition of studying the ways in which the environment has an influence on human health. Epidemiology is the area of 'clinical medicine' that comprises various approaches to the study of these phenomena, by focusing on the relation between outcomes of interest and exposure to the environment, broadly construed to include diet, external chemicals, lifestyle, etc.[1] Philosophical work on epidemiology has increased recently, especially in the context of discussions on causality and causal inference (Broadbent, 2013; Clarke & Russo, 2017; Fuller, 2018). In this paper, I take a step back from the study of products of research, such as causal claims, and turn to the study of the processes, and thus the practices, of epidemiological research.[2] In particular, I investigate the ways in which data is collected, integrated and used as evidence in contemporary epidemiology. Following recent work on data epistemology by Sabina Leonelli, I take a relational approach to data, according to which the evidential and representational value of a dataset is not fixed and predetermined: it is the result of various choices, judgments and assumptions to select data as evidence and order it as representation of phenomena (Leonelli, 2016). This makes the context of data practices, with their specific choices, assumptions, constrains and values, highly significant from an epistemological point of view. In particular, it leaves open questions about the ways in which data practices determine the representational and evidential content of a dataset. In this paper, I argue that one of the ways in which this happens is through a specific type of claims: *evidential claims*. Building on recent scholarship on evidential reasoning in the historical sciences by Alison Wylie and others (Chapman & Wylie, 2016), I view evidential claims as claims that determine the datasets that are to be used as evidence (Sect. 2). I use this notion of evidential claims to specify the epistemic role of data practices and present a typology of methods, approaches and results in contemporary epidemiology. I do so by identifying three main *strategies* for evidential claims. The *macro strategy*, which is implemented at the starting point of research and individuates the dataset that serves as the initial evidential space for research. The micro strategy, which is used to generate evidential claims on the microscopic and individual component of target phenomena. The *association strategy*, that uses evidence from the other strategies to identify a dataset as a representation of the different levels and relations between exposure and disease. As I aim to show, this typology is significant for understanding the current landscape of biomedical research as well as data epistemology (Sect. 4).

I illustrate my analysis with a case study on the epidemiology of the 'exposome'. The exposome was conceived and proposed by Christopher

E-mail address: stefano.canali@philos.uni-hannover.de.

[1] Here I am following a distinction between clinical medicine and bench medicine (Thompson & Upshur, 2017, pp. 2–4; p.24).

[2] On the distinction between scientific products and processes and its importance in the context of the 'practice turn', see Soler and colleagues (Soler, Zwart, Lynch & Israel, 2014, pp. 14–24).

https://doi.org/10.1016/j.shpsc.2019.101248
Received 30 July 2018; Received in revised form 15 October 2019; Accepted 27 December 2019
Available online 16 April 2020

Wild (2005) to describe the totality of exposures to the environment experienced by individuals. Wild introduced the exposome as an umbrella concept, to integrate various ideas and approaches to the study of the relation between exposure and disease (Wild, 2012). Especially after the end of the Human Genome Project, there has been growing consensus on the crucial role played by differences in the environment in the determination of disease (Rappaport & Smith, 2010). Yet, the environment influences health through various pathways, which may involve significantly different phenomena, take place in diverse locations and temporalities, are often difficult to track and measure, and need to be studied at different levels of abstraction. The exposome is presented as a way of overcoming some of these complexities, by considering exposure in its totality.[3] Here, totality is firstly meant to include all the exposures experienced throughout a lifetime. Thus, at any point in an individual's lifetime, their exposome will comprise all the exposures experienced from conception onwards: for instance, the exposome of an adult includes exposures in utero, which may have an impact on health only at a later stage (Robinson & Vrijheid, 2015). Totality is also meant to comprise the substances and processes that an individual is exposed to at a single point in time, both at an external and internal level. In other words, to study an individual exposome requires the study of various levels of investigation, from the macroscopic, external component to microscopic, individual elements (Rappaport, 2011). In my use of exposome research as a case study for data practices in contemporary epidemiology, I follow methodological considerations on case studies as ways of relating specific episodes and cases to certain types of phenomena (Pietsch, 2016). As the phenomenon under study, data practices are particularly interesting in epidemiology: while epidemiologists have traditionally been concerned with the collection and analysis of large datasets (Morabia, 2005), the availability of new, large and varied sources of data is often presented as a significant novelty, especially for issues at the interface of environment and health (Leonelli & Tempini, 2018). This elicits questions on a number of issues, including the ways in which diverse datasets are integrated, given representational content and used as to constitute single bodies of evidence (Leonelli, 2013). In this context, the exposome is an important case to investigate as the innovations and benefits of the approach are indeed often connected to the use of new, large and varied datasets (Fleming et al., 2017, p. 12). My case study of exposome research is empirically grounded in the analysis of the EXPOsOMICS project – a consortium, coordinated by the Imperial College London, that was one of the leading projects on the exposome in Europe between 2012 and 2017 and used the exposome approach for the assessment of disease risk related to air and water pollution on the basis of the integration of various sources of data (Vineis et al., 2017a). The empirical research consisted in: a review of scientific publications, reports and presentations resulting from the project; a research visit to the Department of Epidemiology and Biostatistics of Imperial College London in 2017; and in-depth, semi-structed, qualitative interviews with EXPOsOMICS researchers involved at different levels of the project. The point of conducting empirical research on EXPOsOMICS was to gather information on specific data practices, considering that only little of data practices is directly discussed by researchers in their publications and presentations, and thus following methodological considerations on philosophical studies of scientific practice and data (Boumans & Leonelli, 2013; Leonelli, 2016, pp. 6–9).

2. Evidential claims: lessons from the philosophy of archaeology

The focus of my analysis is data practices, as the context in which relations are established between objects of investigation, data and evidence. Following Leonelli's relational account of data, I view scientific data as objects that can be considered potential evidence for claims about phenomena and circulated among individuals (Leonelli, 2016, Chap.3). This view is in contrast with representational conceptualisations of data, whereby data is taken to be something that represents an aspect of the world in a mind- and context-independent way and has a pre-determined evidential content. Following the relational approach, the evidential value of a dataset rather depends on the ways in which a dataset is used, by whom and for which purposes. As a result, the choices, assumptions, constraints and values involved in data practices are highly significant, as they shape the empirical basis of the process of inquiry. In particular, in the context of data practices, a dataset is given representational content and evidential value; is put in relation to other components of scientific epistemology such as models and theories; and what counts as evidence is determined, differentiated and classified. In this paper, I will focus on the first aspect of the epistemology of data practices and ask the following question: in which ways are datasets given representational and evidential value in the context of data practices?[4] The main argument of the paper is that the representational and evidential value of datasets is determined by what I call evidential claims. In particular, I argue that different types of evidential claims are generated at different stages of research and through different approaches, methods and lines of work. I organise these types in a typology of three strategies for evidential claims, on the basis of my analysis of research in the EXPOsOMICS project.

For the notion of evidential claims, I take inspiration from the philosophy of the historical sciences. Wylie has used the notion to characterise different strategies used by historical scientists to retrieve, contextualise, mobilise and more generally use archaeological data. In Wylie's account, evidential claims are "interpretative hypotheses" and "mediating assumptions", that "establish a link between the material traces that survive archaeologically and the past events or conditions of life that are presumed responsible (in part) for producing these traces" (Wylie, 2000, pp. 231–232). The link established by evidential claims is the product of "a chain of inferences that move from some factual ground to these claims by way of mediating warrants" (Chapman & Wylie, 2016, p. 93). Wylie discusses this chain in terms of "scaffolding" (Wylie, 2017), a notion that has been developed in the context of discussions on cultural evolution by William Wimsatt and James Griesemer (2007). In this literature, scaffolding is defined as the result of "structure-like dynamical interactions with performing individuals that are means through which other structures or competencies are constructed or acquired by individuals or organizations" (Wimsatt, 2013, p. 568). In recent years, the notion has been used by philosophers of science to refer to elements of scientific practice that enable and canalise research. In Wylie's terms, an evidential claim thus consists in the identification of a dataset ("data collected through procedure X") and the specification of its evidential value in the context of the investigation of phenomena ("is evidence of phenomenon Y″) on the basis of the mobilisation of lines of evidence, knowledge and commitments that function as warrants 'scaffolding' the relation between data and phenomena. An example of evidential claim in this account is: "radiocarbon results of ceramics are evidence that the site under investigation was occupied in 1050–1650 BP". Here, the evidential connects a dataset (radiocarbon data) to past phenomena (site occupation in 1050–1650) on the basis of lines of evidence, warrants and backing (e.g. archaeological observations of ceramics and spatial distributions, radiocarbon decay rate and physical chemistry).[5] In this paper, I use the notion of

[3] As noted by referees, the exposome is only one attempt at overcoming these issues. An analysis of the relations with other attempts and approaches in biomedical research can be found in Canali (under review).

[4] This is the main focus of the paper, but I will discuss the role of theoretical and methodological commitments in shaping the use of data and say more on modelling at the end of Sect.4. For more on evidence classification in exposome research, see Canali (2019).

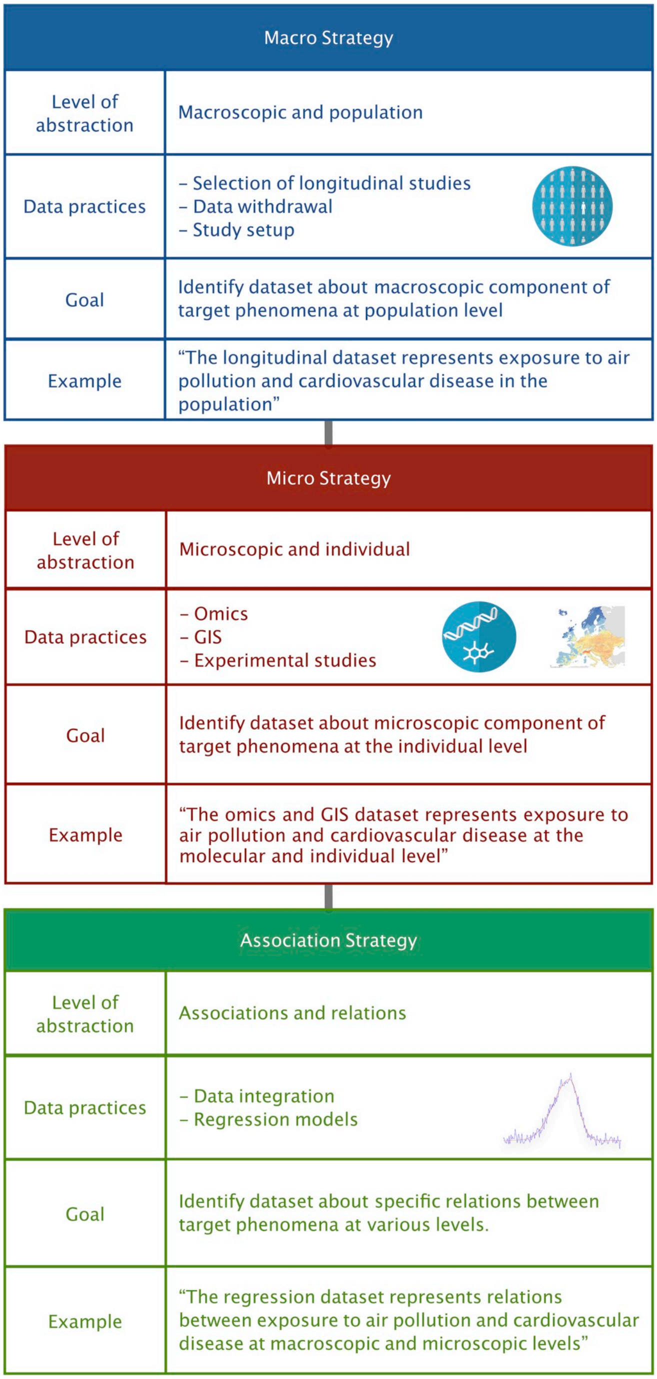
[5] This is a simplified version of the Childers site case analysed by Chapman and Wylie (2016, p.153). I am mostly interested in highlighting the general

evidential claims to characterise data practices as forms of *evidential reasoning*. In other words, I argue that the notion of evidential claims does not only apply to final results or claims, but can also be used to interpret the crucial – if implicit – steps determining the types of phenomena data can be taken to represent and the types of claims data can be evidence for. I use evidential claims as a conceptual tool to unpack the epistemic function of data practices, which I argue is the result of chains of inference based on lines of knowledge and evidence, assumptions, commitments and warrants. For example, the evidential claims that I analyse in the paper include: "this *dataset* is a *representation* of exposure to air pollution and cardiovascular disease in the population", "this *dataset* is a *representation* of exposure to air pollution and cardiovascular disease at the individual level", "this *dataset* is a *representation* of the relations between exposure to air pollution and cardiovascular disease".[6] My use of the concept of evidential claims thus goes in part beyond Wylie's work, and I should also note that in my analysis I take the factual ground of the evidential claims to be datasets and not necessarily material traces. Wylie sometimes uses the word 'data' to describe the factual ground of evidential claims but data is often also used as a synonym to 'evidence', which may lead to some confusion in the context of this paper.[7]

3. Three strategies for evidential claims on the exposome

Using the notion of evidential claims, I now turn to analyse data practices in EXPOsOMICS. I identify the use of three *strategies*, which differ in terms of distinct approaches to the phenomena under study, distinct lines of work that researchers carry out and distinct types of evidential claims. My reconstruction is based on what EXPOsOMICS researchers I interviewed called the "data workflow" of the project. The data workflow was discussed by interviewees as one of the organising principles of the project, which detailed how the different types and journeys of data were organised and managed, how the teams in the project were responsible for different data practices and how the collected data were integrated into singles bodies of evidence for singular studies. I have used this primary classification of data practices in EXPOsOMICS as the starting point of my analysis, as I aim to show that various aspects of data workflows can be interpreted as generating evidential claims (Fig. 1).

I call 'macro' the strategy that is implemented at the starting point of research, as part of the selection and retrieval of samples from longitudinal studies, and individuates the dataset that serves as the initial evidential space for research. This strategy is about the selection of data from longitudinal studies so that it can function as a representation of the target phenomena. The evidential claims from this strategy result in the 'evidential platform' of a study, as the other two strategies work mainly on the data identified by the macro strategy. What I call 'micro strategy' comprises many of the new methods used in exposome research, such as omics. This strategy works on the dataset specified by the macro strategy and is used to perform a microscopic analysis that generates high resolution data on exposure at an internal and external level. The evidential claims of the micro strategy determine the function of data as a representation of the target phenomena at the individual level of participants to a longitudinal study, as opposed the population approach of the macro strategy. I call 'association' the strategy employed when data specified by the micro and macro strategies is statistically analysed and ordered to identify associations between elements and features of the environment and health outcomes. The results of this strategy are evidential claims identifying the integrated dataset as a representation of specific relations between exposure and disease. In the following subsections, I delve into the details of each of these strategies in exposome research and I refer to a study of the relation between air pollution and cardio- and cerebrovascular disease in EXPOsOMICS (Fiorito et al., 2018), as a source of examples to illustrate my argument.



Macro Strategy	
Level of abstraction	Macroscopic and population
Data practices	– Selection of longitudinal studies – Data withdrawal – Study setup
Goal	Identify dataset about macroscopic component of target phenomena at population level
Example	"The longitudinal dataset represents exposure to air pollution and cardiovascular disease in the population"

Micro Strategy	
Level of abstraction	Microscopic and individual
Data practices	– Omics – GIS – Experimental studies
Goal	Identify dataset about microscopic component of target phenomena at the individual level
Example	"The omics and GIS dataset represents exposure to air pollution and cardiovascular disease at the molecular and individual level"

Association Strategy	
Level of abstraction	Associations and relations
Data practices	– Data integration – Regression models
Goal	Identify dataset about specific relations between target phenomena at various levels.
Example	"The regression dataset represents relations between exposure to air pollution and cardiovascular disease at macroscopic and microscopic levels"

Fig. 1. Rendition of the data workflow of EXPOsOMICS in terms of three strategies for evidential claims.

(*footnote continued*)
features of their account here, which is why I am not delving into the specifics of the case.

[6] Here, I am using a broad notion of representation, whereby data can be used in various ways to represent features, elements or structures of target phenomena independently of whether representations mirror real objects. Following Roman Frigg and James Nguyen, "a common misconception needs to be dispelled. The misconception is that a representation is a mirror image, a copy, or an imitation of the thing it represents. On this view representation is *ipso facto* realistic representation. This is a mistake" (Frigg & Nguyen, 2018).

[7] See similar considerations by Bonnin (2019, pp. 3–4) on the difference between 'data' and 'facts'. I will further discuss distinctions between data, representation and evidence at the end of Sect. 3.3 and Sect.4.

3.1. The macro strategy: longitudinal studies

Individual exposome projects focus on specific sources of exposure or types of disease and apply the exposome approach by attempting to study all dimensions of exposure, from the internal to the external level. For example, EXPOsOMICS focused on exposure to air and water pollution and its relation to chronic disease, using data pulled from larger datasets on the basis of longitudinal studies. Longitudinal studies are observational studies that follow a population of interest for an extended period of time. They track variables including clinical states, outcomes of interest, features of participants' lifestyle and their surrounding environment. Data collection is based both on the extraction of physical samples (e.g. blood, cord blood, maternal milk) and questionnaires and one-to-one interviews. Most longitudinal studies data is collected in hospitals by physicians, nurses or trained interviewers. The goal is to bio-monitor and follow participants throughout their life, so as to track the potential development of outcomes of interests. This is usually achieved through 'follow-up', a procedure whereby participants are tracked at different points of their life: they may for instance be asked to answer a new questionnaire years after the initial recruitment, or data may be retrieved through record linkage with hospital, local health authorities and mortality registries. Several longitudinal studies are currently in operation and widely used as sources of evidence in epidemiology. Generally, they tend to differ in terms of: areas covered, as they can be national, subnational, transnational or continental; focus of research, as they may track a variety of disease states and influential factors or be more specifically focused on a handful of phenomena; and time, as some are set up to run constantly and without a predetermined end point, while others come to an end at a specific time but still provide data that can be further analysed. For instance, EXPOsOMICS used the EPIC longitudinal cohort study (the European Prospective Investigation into Cancer and Nutrition), which followed 521,000 participants across 10 European countries for almost 15 year starting in the 1990s, with a broad focus on relationships between diet, nutritional status, lifestyle and environmental factors, and the incidence of cancer and other chronic diseases. But EXPOsOMICS also relied on smaller cohorts such as INMA (INfancia y Medio Ambiente), which is an ongoing study that focuses on Spanish regions and looks at the exposome during pregnancy, in order to investigate the relations between environmental exposures and child development.

A significant portion of epidemiological inquiry is primarily concerned with the establishment and management of longitudinal studies. However, these data collection activities usually precede projects like EXPOsOMICS, where researchers selected and withdrew datasets from specific studies, as opposed to setting up their own. Researchers in EXPOsOMICS want to focus on a specific outcome of interest (such as cardio-vascular disease, child development, increase in oxidative stress, etc.) and, consequently, select a subset of the data collected in the chosen longitudinal study. What I argue is that this selection is based on a type of evidential claim that identifies the chosen dataset as a representation of the specific target phenomena of the study. Following the relational view, material features influence but do not determine the representational value of data; plus, in the case of longitudinal data, the large volume, variety of formats and diversity of types of phenomena to be tracked provide very little constrains on what a subset of data can be about.[8] The representational value of data is therefore to be determined in data practices. I argue that the initial step consists in the selection of a dataset to investigate specific phenomena, such as exposure to air pollution, and that this selection is based on implicit evidential claims that restrict the content of a dataset as a representation of phenomena. In Wylie's terms, these evidential claims function as interpretative hypotheses about the extent to which the longitudinal dataset can be used to represent the target phenomena. This result is obtained through a chain of inference which involves warrants such as the mobilisation of other lines of evidence (e.g. on the setup and work of the longitudinal cohort study), the use of background knowledge (e.g. on disease aetiology) and the grounding on assumptions and commitments (e.g. on data comparability and integration). For example, in the context of the study by Fiorito and colleagues (Fiorito et al., 2018), the initial step of their study was the withdrawal of data from the Italian sub-cohort of EPIC. This consisted in the setup of a case-control study, i.e. the identification of incident and control cases of cardio- and cerebrovascular disease that arose during the follow-up period, with a total of 18,982 individuals identified (Fiorito et al., 2018, p. 236). The identification of this dataset can thus be interpreted as the result of an evidential claim, such as "the dataset about 18,982 participants represents exposure to air pollution and cardiovascular disease in the population". As such, the evidential claim identifies an evidential space, in which a representational relation can be established between the dataset about 18,982 participants, pulled from the EPIC cohort, and the target phenomena under study, i.e. exposure to air pollution and cardiovascular disease. The claim, in turn, is based on warrants including other lines of evidence, knowledge and assumptions. For instance, Fiorito and colleagues mentioned existing evidence about relations between air pollution and coronary events, such as the results of a meta-analysis associating increase in exposure to particulate air matter with an increased risk of coronary events (Fiorito et al., 2018, p. 235). They also relied on background knowledge on a number of mechanistic explanations that have been proposed to connect air pollution to cardiovascular disease. On this basis, they identified a dataset about cases of cardiovascular disease as a potentially interesting evidential space to investigate relations between air pollution and cardiovascular disease, i.e. the target phenomena of their study. This step was also made possible by assumptions and commitments, which for instance included the internal comparability and generalisability of the longitudinal dataset, as these studies collect data at quite different points in time and space and involving diverse types of expertise, lines of work and background (health professionals in hospitals, epidemiologists setting up and coordinating the cohort, data analysts in projects like EXPOsOMICS, etc.). By selecting a subset of data, researchers also committed to a certain level of generalisability. Namely, longitudinal cohort studies data is local by default, having been collected in person and/or physically extracted from participants, and this locality is important for projects like EXPOsOMICS, which aimed to connect features of specific kinds of environment with specific health states. Yet, a dialogue between locality and generalisability was stressed multiple times in EXPOsOMICS, as researchers aimed at finding results that were sensible to local features but also general enough for similar populations. More broadly, the evidential claims of the macro strategy are ways of committing to a certain interpretation of the longitudinal dataset as a representation of the target phenomena. The evidential space established by this type of evidential claims brings these commitments to bear on the kind of research that is carried out at later stages of the project.

3.2. The micro strategy: omics, GIS and experimental studies

I have mentioned that one of the ways in which the exposome approach is considered innovative is the study of both the internal and external component of the exposome. This is connected to what I call the micro strategy, that is focused at a lower level of investigation than the macro strategy and is aimed at producing evidential claims at the microscopic level of investigation. The micro strategy is applied at different points of exposome research and operates to generate evidential claims at the individual level of both the internal molecular

[8] See Leonelli (2019, pp. 17–19) for an extensive discussion of the relation between material features and the representational value of data according to the relational view.

environment and the external surrounding environment.[9] This strategy is thus particularly complex, requires significantly different lines of work and generates significantly different types of data, and its role is to establish a representational relation that connects this diverse data to the target phenomena at the individual level. A first step where this approach is applied in exposome research is omics. Omics are analytical methods and techniques that have been developed in genomic and sequencing projects to quantify and study molecules and various processes in the cell, such as proteins, metabolism, DNA adducts, epigenetic changes and mRNA expressions. In the context of exposome research, the use of these techniques is connected to the idea of the exposome as the totality of exposure, whereby the exposome comprises elements, substances and processes also at internal levels. Namely, exposure to e.g. air pollution can cause the presence of or reaction to toxicants at the molecular level, which in turn can yield disease. Exposome researchers use omics to study the internal component of exposure, in order to look for traces of external exposures and potential initial reactions at the molecular level. Omics measurements are performed using methods including mass spectrometry, whose results are visualised in terms of plots with lines and peaks; these peaks – also known as features – are indicative of the chemical composition of the sample. The results obtained through omic analysis of the samples are, thus, a list of molecules per each sample. In exposome research, these are used to get a picture of the molecular composition of the samples and thus try to understand the potential influence of pollutants on the internal, molecular component of the exposome. The resulting tables of data are usually called exposure or omic profiles; their type depends on the omics technique used for the analysis, which in turn depends on the kind of molecular features researchers intend to focus on for their particular project and on the nature of the samples or the process they want to study.[10] For example, in the study carried out by Fiorito and colleagues, EXPOsOMICS researchers used proteomics, an omics technique for the study of the proteins produced in an organism, as a way to collect data on inflammation-related proteins. One of the most significant consequences of the study of the internal component of the exposome through omics data is that it has elicited the collection of data at a similar level of abstraction and resolution for the external component of the exposome. In EXPOsOMICS, this line of work within was carried out by the 'GIS team', whose task was to develop Geographical Information Systems (GIS) to visualise, model and analyse geographical data. These are used to generate individual estimates of the chemicals and pollutants that could have been present in the environment and to which each participant could have been exposed. On the basis of the postcodes of the areas where participants lived during the cohort study period, in EXPOsOMICS the GIS team retrieved other data, from maps of characteristics of populations that are routinely collected by monitoring stations, which are usually located at various city locations in Europe and provide information on the conditions and features of the area. All these variables were used to tweak a geo-spatial model which, taking into account these variables and differences in populations, assigned an estimate of the presence of toxicants to which every participant could have been exposed during the study (Gulliver et al., 2018). For instance, in the case of Fiorito and colleagues' study, the GIS team generated estimates of air pollution concentrations at the postcode areas of the participants to the study, including concentrations of NO_2, NO_x and $PM_{2.5}$.

What I want to highlight in this context is the goal of representing target phenomena at a microscopic and individual level, to be added to the picture developed through the macro strategy. What I call micro strategy is thus used to further specify the evidential and representational content of data. I argue that this is the result of a specific type of implicit evidential claims, that interpret data generated through omics and GIS as a representation of the target phenomena. In the case Fiorito and colleagues, the micro strategy was used to individuate a proteomic dataset that could be used as a representation of internal exposure and a NO_2, NO_x and $PM_{2.5}$ dataset that could be used as a representation of external exposure, both at the individual level. The evidential claim integrated these datasets to determine a single representation of target phenomena, being of the kind "The dataset comprising proteomics and NO_2, NO_x and $PM_{2.5}$ represents internal and external exposure at the level of individual participants". This representational relation is the result of a particularly complex chain of inferences, based on warrants including various lines of evidence, background knowledge and material components of the project. For example, Fiorito and colleagues relied on other lines of evidence, such as routine background monitoring to back-extrapolate GIS estimates, background knowledge from molecular biology about proteins and inflammation, as well as the techniques and infrastructures used to implement both GIS and omic analyses. I want to highlight the very important role that assumptions and commitments play here, especially in the context of the comparability and 'mirroring' of omics and GIS data. The epistemic role of GIS data practices in exposome research is to get a level of detail and resolution which is on par with omics data. In a project like EXPOsOMICS, the work of the GIS team was not only to give more detail to available data on external exposure, but also to level and balance data on external and internal exposure. The GIS data identified by the strategy consisted in tables of data with estimates for different chemical compounds assigned to the members of the cohort, thus to an extent mirroring the exposure profiles produced by omics. Yet, GIS data remained substantially different from omics data, in terms of the methods used to generate the data, the uses of the results and the types of shared knowledge, theories and methods. Omics profiles provided the structure of various molecules under study, while GIS data provided estimates of the concentrations of single pollutants. On the one hand, omics data was the result of the analysis of physical samples, while on the other hand GIS data was collected through back-extrapolated estimates. Omics tables are very wide, whilst GIS data on environmental factors is just one column wide. In addition, in EXPOsOMICS omics data was usually collected by epidemiologists working at the core of the research team of the project, whilst GIS is a sub-discipline of information systems. Hence, how can these diverse datasets be used as a single body of evidence about target phenomena at the individual level? While this is a problem for the representational view, according to which the material and physical features of the data fix its representational value, following a relational approach these features do not uniquely cause what the data is about. Rather, the "aboutness" of the data is determined (also) by specific choices and judgements at the level of data practices. My analysis specifies the role of data practices in this context, as I argue that the use of omic and GIS data as an integrated dataset is based on an implicit evidential claim, that on the basis of the series of aforementioned warrants poses an interpretation of omics and GIS data as representing exposure at the internal and external level.[11] In addition, in some areas of exposome research the strategy also involves experimental studies. For instance, in EXPOsOMICS personal monitoring and tracking devices were used to perform real-time measurements of exposure levels and physiological variables in controlled environments (Vineis et al., 2017a, pp. 148–149). These studies included tracking

[9] A referee noted that 'environment' suggests external exposure in epidemiology. The view of the body and the internal, molecular level as an 'environment' is indeed one of the peculiar features of the exposome approach. For more on conceptual changes related to the exposome and their relation to the notion of environment, see Canali (under review).

[10] In my interviews, EXPOsOMICS researchers discussed data produced through omic analyses as "big data". The notion of big data was here connected to the possibility of getting a large volume of data for a small number of individuals, in contrast to more traditional scenarios where many individuals are studied through a smaller number of variables.

[11] I shall come back to these commitments and their importance in debates on contemporary biomedical research in the next section.

participants while they were, for example, walking in areas with clearly contrasting pollution levels, such as Oxford Street and Hyde Park in London (Espín-Pérez et al., 2018). In these studies, the same participants experienced several different conditions, which allowed EXPOsOMICS researchers to look at high effects of exposure, as opposed to long-term effects that are tracked through longitudinal studies. Finally, one way of interpreting the evidential claims of this strategy could be in causal terms, in the sense that, to an extent, these evidential claims can be used to later identify causal relations in the complicated web of relations between the presence of pollutants in the environment, exposure, reactions to exposure and (potentially) development of disease. In EXPOsOMICS, the theoretical background of the use of molecular data was usually discussed in terms of a methodological approach known as the "meet in the middle approach" (Chadeau-Hyam et al., 2011), whose goal is to investigate what lies 'in the middle' of the associations between exposure and outcomes of interest and therefore test the causal nature of statistical associations.[12]

3.3. *The association strategy: data integration and evidence production*

One of the aims of exposome studies is to provide evidence about the assessment of certain exposures: the final specification of this type of evidence – I argue – is the result of evidential claims developed by what I call association strategy. This strategy is based on an approach focused on associations between exposure and outcomes of interest and is an integration of the data identified through the macro and micro strategies. The dataset specified by these strategies is used in a regression model, looking for associations between exposure and health outcomes. In EXPOsOMICS, the regression model looked at one exposure profile and one omics feature at a time. This means that, for example, each of the 4000 features that can be measured in metabolomics was modelled in association with data on external exposure.[13] Then, data analysts in EXPOsOMICS looked at the models where the association was statistically significant, after taking into account that, whilst doing thousands of tests, some will be statistically significant only by chance. On this basis, EXPOsOMICS researchers for instance found that about 170 out of the 4000 features were associated with a specific outcome.[14] The regression models integrated data specified by the previous strategies and gave it a specific order as a representation of the relations between target phenomena at different levels (internal and external, population and individual). I argue that this ordering is the result of the interpretative commitment of evidential claims that specify the evidential value of the resulting dataset as representing the phenomena. This is a result of a long chain of inference, based on warrants including the various lines of evidence specified in previous strategies, sources of knowledge mobilised, types of expertise and material components of the project and commitments and assumptions. The very use as a single body of evidence of such a diverse dataset in terms of type, format, resolution, volume etc. is the result of commitments, and in particular a commitment to the validity of using data collected with both traditional (longitudinal studies) and innovative methods (omics and GIS), which is a specificity of the exposome approach and raises a number of questions about the role of the exposome in contemporary biomedical research (see Sect. 4). The regression models used at this stage are also based on more specific assumptions, such as that the prediction of the outcome of interest is a matter of multiple factors. For example, when using omics data, models were based on the assumption that a pool of different omic features predict the outcome of interest better than each of them separately or their sum. These models are thus not primarily causal: their goal is prediction, and more specifically exposome researchers are after the model that best predicts the outcomes of interest, taking into consideration the omics variables jointly, as opposed to using one by one or summing the predictions separately. The models were adjusted on the basis of the data generated at the individual level. For example, the data produced through the micro strategy and experimental studies was used to try and see, when the exposure is modelled, how well the actual exposure a participant was exposed to can be predicted, to try to make sure that what is modelled is in the range of what is measured.

The result of this strategy is thus the development of an implicit evidential claim, which in the case of Fiorito and colleagues' study was of this kind: "The integrated dataset represents the different levels and relations of exposure to air pollution and cardiovascular disease". As the endpoint of the chain of evidential claims I have discussed, the association strategy plays the distinct role of elaborating what in the intentions of exposome researchers is a potentially full representation of phenomena, which in turn characterises the final dataset as evidence that can be used in the context of other types of claims, such as knowledge claims about specific relations between exposure and disease. For instance, as a result of their study Fiorito and colleagues claimed that their work is evidence of a strong association between air pollution and cardiovascular disease, mediated by oxidative stress and inflammation. I interpret this final step of their research as a knowledge claim about target phenomena, that is built on the specification of their dataset as a representation of the phenomena under study by evidential claims. In other words, according to my typology the final result of an exposome study is based on three strategies, which function as part of data practices applied at various levels, and on different datasets to specify and identify the evidential content and basis of the study.

4. Data practices, epistemic strategies and the role of evidential claims

The way in which I have structured the previous section and I refer to examples from research in EXPOsOMICS might suggest that the strategies are linear and subsequently ordered. However, this is not necessarily the case: the strategies are interrelated in both subsequent and non-linear ways, through at least two different types of relations (Fig. 2). The first is straightforward: the evidential claims generated through the macro and micro strategy provide data that is used as the evidential basis for the micro and association strategy (see the central arrows). For example, as we have seen in the previous section, the macro strategy was employed by Fiorito and colleagues to identify data on cases and controls of cardio- and cerebrovascular disease within the EPIC cohort study. This was then used as the basis of the micro strategy, where the dataset was analysed to specify a different component of representation. Yet, the macro and micro strategies are also individually connected to the association strategy through a relation of specification, as the epistemic commitments and assumptions that establish the evidential relations in the macro and micro strategies limit the kind of the evidential claims that can be generated through the association strategy (see the dashed arrows). As we have seen, the evidential claims of the macro and micro strategies provide a constrain to the generalisability of the evidential claims of the association strategy. In the study of Fiorito and colleagues, the identification of the dataset on cases of cardio- and cerebrovascular disease through the macro strategy specified volume (18,982 individuals) and type of the study (case-control), which in turn limited the extent to which the statistical results of the association strategy can be applied. Similarly, the commitments of the micro strategy, most importantly the assumptions on the comparability of omics and GIS datasets, provide a constrain to how generalisable the results of the association strategy are.

[12] Vineis (2015) argues that the meet in the middle approach used in EXPOsOMICS is a "very rudimentary approach to causality", but links it to the work of Wesley Salmon on the propagations of marks and causal processes (Vinies, 2015, p. 720).

[13] Metabolomics is an omic technique used for the study of the processes and products of metabolism.

[14] These models were usually developed through a mix of univariate and multivariate methods. For an overview of the methods used in the project and for the study of the exposome, see Chadeau-Hyam et al. (2013).

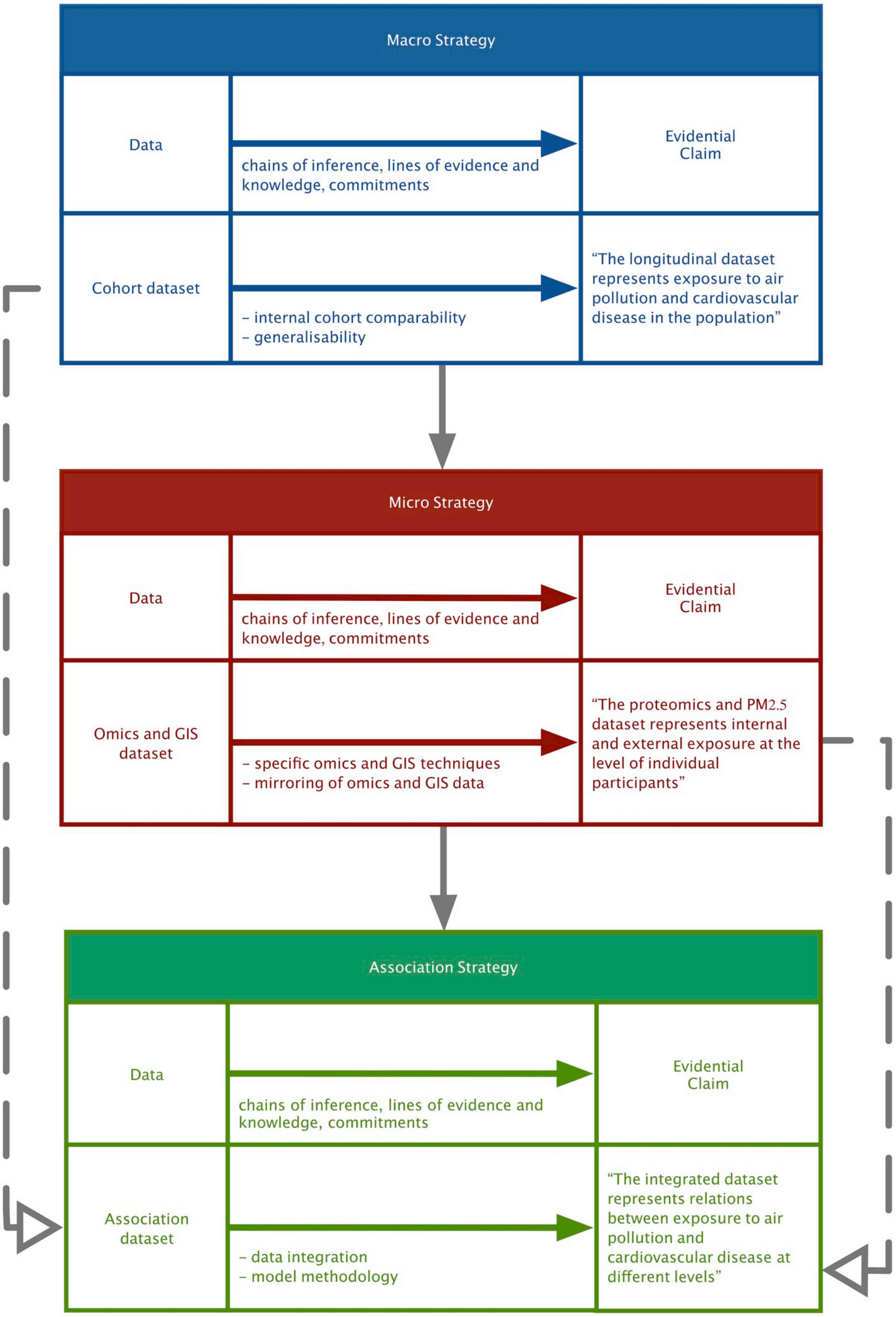

Macro Strategy
Data
chains of inference, lines of evidence and knowledge, commitments
Evidential Claim
Cohort dataset
– internal cohort comparability
– generalisability
"The longitudinal dataset represents exposure to air pollution and cardiovascular disease in the population"
Micro Strategy
Data
chains of inference, lines of evidence and knowledge, commitments
Evidential Claim
Omics and GIS dataset
– specific omics and GIS techniques
– mirroring of omics and GIS data
"The proteomics and PM2.5 dataset represents internal and external exposure at the level of individual participants"
Association Strategy
Data
chains of inference, lines of evidence and knowledge, commitments
Evidential Claim
Association dataset
– data integration
– model methodology
"The integrated dataset represents relations between exposure to air pollution and cardiovascular disease at different levels"

Fig. 2. Reconstruction of the relations between the macro, micro and association strategies.

An analysis of these strategies and their interrelations in exposome research is connected to discussions on and allows for comparisons with other areas of contemporary biomedical research.[15] In this sense, what I have defined as the macro strategy is a relatively traditional way of defining the evidential space of longitudinal data and is used throughout epidemiology (Morabia, 2005). Similarly, while the specific features of the methods used as part of this strategies may vary, the general approach of the association strategy can be considered typical of epidemiology more generally, as the discipline is usually discussed as the study of the associations and distributions of health and disease in a population (Broadbent, 2013). Conversely, the micro strategy, with the use of omics and GIS data, is one of the more innovative and peculiar aspects of exposome research. More and more areas of epidemiology are trying to make use of the increasingly available data collected at the micro level, which creates potential benefits but also poses significant challenges on epistemic, material and organisational components of research (Fleming et al., 2017, p. 10). For example, Leonelli and Tempini have analysed 'data mash-ups' combining data from epidemiology, biomedicine and climate and environmental science and have argued that the use of these diverse datasets is made possible by specific invariant strategies, which in turn presents a number of challenges (Leonelli & Tempini, 2018). My discussion of the micro strategy in exposome research is in line with these considerations, as I have argued that the strategy is based on various and significant commitments and assumptions about omics and GIS data. This discussion is also

[15] This aligns my case study on the exposome with other considerations on case studies by Mary Morgan, who argues that comparisons and bridges with other contexts are among the use-values of this approach (Morgan, 2018).

connected with broader scientific and philosophical debates on contemporary biomedicine, on issues such as the raise and expansion of molecular approaches in medical research (Boniolo & Nathan, 2017), the relations and differences between genomics and 'postgenomics' (Stevens & Richardson, 2015) and the expansion of genomic-based, data-intensive techniques in other fields (Leonelli, 2018). In this context, my analysis of omics data in exposome research shows how its employment in the epidemiological context crucially rests upon a number of assumptions and commitments. In particular, I have emphasised how the use of omics plays a significant role at different steps of exposome research, as it: requires epistemic, material and organisational components that are new to epidemiology; elicits the use of other, new sources of data such as GIS; and has an influence on the methods and techniques applied at the level of the association strategy. My analysis thus specifies and delimits the epistemic value of genomic-based techniques and data in epidemiology. At the same time, it brings into question the use of the micro strategy alongside more traditional epidemiological approaches like the macro strategy and the extent to which the exposome goes beyond genomic-based approaches and the genome (Canali, under review). In connection with other analyses of 'micro strategies', the typology provides a platform for investigating which challenges and benefits are linked to specific approaches such as the exposome and why some approaches are considered better than others in biomedical research.

At the same time, investigating these strategies is way of studying data epistemology and the role of data and evidence in the process of inquiry. In Wylie's view, evidential claims relate material traces to phenomena of the past. Here, I use this notion to interpret data practices in contemporary epidemiology as developing claims that identify datasets as specific representations. This use of evidential claims thus points to a view of evidence as the product of various epistemic practices, claims, and considerations – as a category that identifies specific data that is used for specific purposes. In this sense, evidence is not to be considered a necessarily fixed and stable entity, and neither a synonym of data. Rather, identifying a dataset as evidence is a way of expressing claims about phenomena and, potentially, knowledge on the world. In this context, Leonelli (2019) has recently given an account of data practices that distinguishes between activities that define the evidential space of data and make it usable as evidence (data processing) and activities aimed at the ordering and clustering of data to represent phenomena (data modelling). Her distinction runs parallel to my analysis of the ways in which representational and evidential content is given to data in exposome research on the basis of various judgements. Following Leonelli's account, the macro and the micro strategies could be considered instances of data processing, as they restrict the evidential space of longitudinal data, and the association strategy could be interpreted as data modelling, in the sense of clustering data from the previous strategies to represent the target phenomena. At the same time, the case I have looked at shows how the context of data processing and the context of data modelling are sometimes difficult to neatly separate: for example, the micro strategy is as much about making data usable as evidence as it is about developing a specific component and type of representation. Indeed, in Leonelli's analysis of data practices the two steps are often intertwined and many choices and judgements at the level of data processing have an influence on data modelling.[16] Furthermore, Leonelli's account complements the ways in which I have differentiated between the use of data for representation and for evidence. Namely, the representational value of data is established by various practices employed by researchers, which in turn determine how and for what data can be used as evidence. Yet, the step in which the final dataset is used as evidence is consequent and, in particular, data acquires the status of evidence at the point in which it is used in specific knowledge claims.[17]

5. Conclusions

How is the representational and evidential value of a dataset determined in scientific practice? In this paper, I have provided an answer to this question in the context of research on the epidemiology of the exposome. I have argued that the representational and evidential content of data is determined by evidential claims, which delimit the evidential space of data on the basis of chains of inference, the mobilisation of knowledge and the production of assumptions, commitments and warrants. This has led me to distinguish between three strategies for different types of evidential claims: the macro strategy, which generates claims that restrict samples and provide the initial evidential space for research; the micro strategy, which specifies a representation of target phenomena from the perspective of the individual level of participants; and the association strategy, which provides evidential claims to establish the dataset to be used in knowledge claims about associations between exposure and disease. Choosing data as units of philosophical analysis, I have shown the epistemic function of data practices and characterised them as forms of evidential reasoning. I should note that this focus is different from most philosophical analyses of epidemiology, which have often studied epidemiological research in causal terms. This work on causation has had the merit of putting epidemiology on the map of disciplines of interest to philosophy of science. Additionally, it has led philosophers to study epidemiological practice and methodology and collaborate with epidemiologists, directly engaging in current discussions in the scientific literature – a very significant result.[18] Plus, this focus is coherent with the 'end goal' of epidemiology, that is intervening on populations to improve public health. I do not intend my analysis to be in contrast with this line of research, but rather to complement it. A philosophy of epidemiology concerned with causality will mostly focus on the results of epidemiological research, and may thus end up overlooking the epistemic role of processes and elements that proceed final results but significantly influence them. At the same time, I want to note that my typology is not intended to be exhaustive of all these elements and the ways in which epidemiological research can be performed. For once, while in the cases I have analysed the three strategies are used alongside each other, other projects might only use one of them and/or together with other approaches. For instance, the macro strategy is used as a starting point of many projects, as a way of defining the evidential space of longitudinal data, which might then be specified in different ways than the micro and association strategy discussed here. At the same time, some of the components and details of the strategies may vary: for example, the methods employed as part of the association strategy might be significantly different than those discussed here, depending on the specifics of the study and the area where the strategy is used. Despite these complexities, I hope to have shown the merit of distinguishing 'local' strategies in data practices, as a way to individuate epistemic processes at the interface between interactions with the world, data and evidence; and to specify how the epistemic value of datasets lies in the relations established between data and knowledge, claims, models, commitments, theories – rather than in the data itself, however big, varied or generated at a fast rate.

[16] See e.g. "both data processing and data ordering contribute to carving out what phenomena researchers are actually able to produce knowledge about" (Leonelli, 2019, p. 20).

[17] Here, I am using a broad meaning of 'knowledge', which in turn leaves open questions about, for instance, what characterises these knowledge claims and whether they are evidential claims of the same type as those I have discussed throughout the paper. Answering this question is a necessary task, yet one I do not have space to address here.

[18] See for instance the work of philosophers Federica Russo and Phyllis Illari in collaboration with exposome epidemiologist Paolo Vineis (Vineis, Illari, & Russo, 2017b) and philosopher Alex Broadbent with various colleagues (Vandenbroucke, Broadbent, & Pearce, 2016).

Acknowledgments

I first wish to thank all the EXPOsOMICS researchers who shared their time and facilities with me to discuss their fascinating work. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project 254954344/GRK2073, as part of the graduate research group "Integrating Ethics and Epistemology of Scientific Research" (https://grk2073.org). I am deeply grateful to this group for all the support, discussions and feedback; special thanks to David Hopf and Saana Jukola for reading a draft of this article. I am also indebted to the hospitality of Egenis, the Centre for the Study of the Life Sciences, at the University of Exeter, and particularly the Exeter Data Studies group (https://datastudies.eu), where I started developing the ideas of this paper and the empirical research on EXPOsOMICS. Warm thanks to friends and mentors: Thomas Bonnin, Annamaria Carusi, Sabina Leonelli, Guglielmo Militello, Thomas Reydon, Federica Russo and to three anonymous referees for closely reading and commenting the article. Finally, thanks for the feedback to participants to the ISHPSSB session "Challenges and Opportunities of Data Integration" (São Paolo, July 2017) and the PSA2018 poster session (Seattle, November 2018), where I presented previous drafts of this paper.

References

Boniolo, G., & Nathan, M. J. (Eds.). (2017). *Philosophy of molecular medicine: Foundational issues in research and practice*. London, UK: Routledge.

Bonnin, T. (2019). Evidential reasoning in historical sciences: Applying Toulmin schemes to the case of Archezoa. *Biology & Philosophy, 34*(30), https://doi.org/10.1007/s10539-019-9677-z.

Boumans, M., & Leonelli, S. (2013). Introduction: On the philosophy of science in practice. *Journal for General Philosophy of Science, 44*(2), 259–261.

Broadbent, A. (2013). *Philosophy of epidemiology*. Basingstoke: Palgrave Macmillan.

Canali (under review). The exposome as a postgenomic repertoire: Exploring scientific change in contemporary epidemiology.

Canali (2019). Evaluating evidential pluralism in epidemiology: Mechanistic evidence in exposome research. *History & Philosophy of the Life Sciences, 41*(1), 4.

Chadeau-Hyam, M., Athersuch, T. J., Keun, H. C., De Iorio, M., Ebbels, T. M. D., Jenab, M., et al. (2011). Meeting-in-the-middle using metabolic profiling - a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers: Biochemical Indicators of Exposure, Response, and Susceptibility to Chemicals, 16*(1), 83–88.

Chadeau-Hyam, M., Campanella, G., Jombart, T., Bottolo, L., Portengen, L., Vineis, P., et al. (2013). Deciphering the complex: Methodological overview of statistical models to derive OMICS-based biomarkers. *Environmental and Molecular Mutagenesis, 54*(7), 542–557.

Chapman, R., & Wylie, A. (2016). *Evidential reasoning in archaeology*. London, UK: Bloomsbury Academic Publishing.

Clarke, B., & Russo, F. (2017). Causation in medicine. In J. A. Marcum (Ed.). *The bloomsbury companion to contemporary philosophy of medicine* (pp. 297–322). London, UK: Bloomsbury.

Espín-Pérez, A., Font-Ribera, L., van Veldhoven, K., Krauskopf, J., Portengen, L., Chadeau-Hyam, M., et al. (2018). Blood transcriptional and microRNA responses to short-term exposure to disinfection by-products in a swimming pool. *Environment International, 110*, 42–50.

Fiorito, G., Vlaanderen, J., Polidoro, S., Gulliver, J., Galassi, C., Ranzi, A., et al. (2018). Oxidative stress and inflammation mediate the effect of air pollution on cardio- and cerebrovascular disease: A prospective study in nonsmokers. *Environmental and Molecular Mutagenesis, 59*(3), 234–246.

Fleming, L., Kessel, A., Murray, V., Depledge, M., Leonelli, S., Tempini, N., et al. (2017). *Big data in environment and human health. Oxford research encyclopedia of environmental science*. Oxford (UK): Oxford University Press.

Frigg, R. P., & Nguyen, J. (2018). Scientific representation. In E. N. Zalta (Ed.). *The stanford encyclopedia of philosophy*. (Winter 2018 edition) https://plato.stanford.edu/archives/win2016/entries/scientific-representation, Accessed date: 15 October 2019.

Fuller, J. (2018). Universal etiology, multifactorial diseases and the constitutive model of disease classification. *Studies in History and Philosophy of Biological and Biomedical Sciences, 67*, 8–15.

Gulliver, J., Morley, D., Dunster, C., McCrea, A., van Nunen, E., Tsai, M.-Y., et al. (2018). Land use regression models for the oxidative potential of fine particles (PM 2.5) in five European areas. *Environmental Research, 160*, 247–255.

Leonelli, S. (2013). Integrating data to acquire new knowledge: Three modes of integration in plant science. *Studies in History and Philosophy of Biological and Biomedical Sciences, 44*(4), 503–514.

Leonelli, S. (2016). *Data-centric biology: A philosophical study*. Chicago, IL: University of Chicago Press.

Leonelli, S. (2018). Introduction to Genomics and DNA-based technologies in the clinic and beyond. In S. Gibbon, B. Prainsack, S. Hilgartner, & J. Lamoreaux (Eds.). *Handbook of genomics, health and society* (pp. 11–14). (2nd ed.). London: Routledge.

Leonelli, S. (2019). What distinguishes data from models? *European Journal for Philosophy of Science, 9*(2), 22.

Leonelli, S., & Tempini, N. (2018). Where health and environment meet: The use of invariant parameters in big data analysis. In S. Valles, & J. Kaplan (Eds.). *Synthese (online first), special issue "Philosophy of Epidemiology"*.

Morabia, A. (2005). Epidemiology: An epistemological perspective. In A. Morabia (Ed.). *A history of epidemiologic methods and concepts* (pp. 3–125). Berlin: Springer.

Morgan, M. S. (2018). *Exemplification and the use-values of cases and case studies. Studies in history and philosophy of science: Part A.* (online first).

Pietsch, W. (2016). Two modes of reasoning with case studies. In T. Sauer, & R. Scholl (Eds.). *Boston studies in the philosophy and history of science*. (pp. 49–67). Springer.

Rappaport, S. M. (2011). Implications of the exposome for exposure science. *Journal of Exposure Science and Environmental Epidemiology, 21*(1), 5–9.

Rappaport, S. M., & Smith, M. T. (2010). Exposome, environment and disease risks. *Science, 330*(6003), 460–461.

Robinson, O., & Vrijheid, M. (2015). The pregnancy exposome. *Current Environmental Health Reports, 2*(2), 204–213.

Soler, L., Zwart, S., Lynch, M., & Israel-Jost, V. (Eds.). (2014). *Science after the practice turn in the philosophy, history, and social studies of science*. London: Routledge.

Stevens, H., & Richardson, S. S. (2015). Approaching postgenomics. In S. S. Richardson, & H. Stevens (Eds.). *Postgenomics: Perspectives on biology after the genome* (pp. 1–8). Durham: Duke University Press.

Thompson, R. P., & Upshur, R. (2017). *Philosophy of medicine: An introduction*. London, UK: Routledge.

Vandenbroucke, J. P., Broadbent, A., & Pearce, N. (2016). Causality and causal inference in epidemiology: The need for a pluralistic approach. *International Journal of Epidemiology, 45*(6), 1776–1786.

Vineis, P. (2015). Exposomics: Mathematics meets biology. *Mutagenesis, 30*(6), 719–722.

Vineis, P., Chadeau-Hyam, M., Gmuender, H., Gulliver, J., Herceg, Z., Kleinjans, J., et al. (2017a). The exposome in practice: Design of the EXPOsOMICS project. *International Journal of Hygiene and Environmental Health, 220*(2), 142–151.

Vineis, P., Illari, P., & Russo, F. (2017b). Causality in cancer research: A journey through models in molecular epidemiology and their philosophical interpretation. *Emerging Themes in Epidemiology, 14*(1), 7.

Wild, C. P. (2005). Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology Biomarkers and Prevention, 14*(8), 1847–1850.

Wild, C. P. (2012). The exposome: From concept to utility. *International Journal of Epidemiology, 41*(1), 24–32.

Wimsatt, W. C. (2013). Articulating Babel: An approach to cultural evolution. *Studies in History and Philosophy of Biological and Biomedical Sciences, 44*(4), 563–571.

Wimsatt, W. C., & Griesemer, J. R. (2007). Reproduction entrenchments to scaffold culture: The central role of development in cultural evolution. In R. Sansom, & R. N. Brandon (Eds.). *Integrating evolution and development: From theory to practice* (pp. 227–324). MIT Press.

Wylie, A. (2000). Questions of evidence, legitimacy, and the (Dis)Unity of science. *American Antiquity, 65*, 227–237 02.

Wylie, A. (2017). How archaeological evidence bites back. *Science, Technology & Human Values, 42*(2), 203–225.