Assessing correlations between geological hazards and health outcomes: Addressing complexity in medical geology

Nicola Ann Wardrop a,1, Jennifer Susan Le Blond b

a Geography and Environment, University of Southampton, Southampton, UK
b Department of Earth Science and Engineering, Imperial College London, London, UK

A R T I C L E   I N F O

Article history:
Received 26 May 2015
Received in revised form 9 July 2015
Accepted 20 July 2015
Available online xxxx

Keywords:
Medical geology
Environmental epidemiology
Exposure assessment

A B S T R A C T

Background: The field of medical geology addresses the relationships between exposure to specific geological characteristics and the development of a range of health problems: for example, long-term exposure to arsenic in drinking water can result in the development of skin conditions and cancers. While these relationships are well characterised for some examples, in others there is a lack of understanding of the specific geological component(s) triggering disease onset, necessitating further research.

Objectives: This paper aims to highlight several important complexities in geological exposures and the development of related diseases that can create difficulties in the linkage of exposure and health outcome data. Several suggested approaches to deal with these complexities are also suggested.

Discussion: Long-term exposure and lengthy latent periods are common characteristics of many diseases related to geological hazards. In combination with long- or short-distance migrations over an individual’s life, daily or weekly movement patterns and small-scale spatial heterogeneity in geological characteristics, it becomes problematic to appropriately assign exposure measurements to individuals. The inclusion of supplementary methods, such as questionnaires, movement diaries or Global Positioning System (GPS) trackers can support medical geology studies by providing evidence for the most appropriate exposure measurement locations.

Conclusions: The complex and lengthy exposure–response pathways involved, small-distance spatial heterogeneity in environmental components and a range of other issues mean that interdisciplinary approaches to medical geology studies are necessary to provide robust evidence.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The geological characteristics of the earth’s surface can directly influence human health via the ingestion, inhalation or absorption of specific elements or compounds derived from naturally occurring materials (e.g. Davies et al., 2004; Skinner, 2007). The degree to which we understand the relationship between exposure and health outcomes, however, varies significantly between different geological hazards within the environment. For example, the relationship between exposure to water and food supplies contaminated with arsenic, and the development of skin conditions and a variety of cancers is well known (Bhattacharya et al., 2012; Naujokas et al., 2013). However, while the association between specific soil types and the development of podocnosis (non-infectious elephantiasis) has been established, the specific components within the soil that may trigger the onset of podocnosis have not yet been identified (Molla et al., 2014). When considering the discrepancies in our understanding of geological hazards, there are a number of important issues that must be addressed to enable us to explore the relationship between the environment and human health, most notably the compatibility between data collected to determine the potential hazard within the environment and that gathered to estimate disease occurrence.

Using statistical methods to link epidemiological data with geological characterisations can provide improved understanding of the etiologies of environmental diseases, but this linkage is not a straightforward one. Using a range of examples from medical geology, this paper aims to highlight several important complexities that need to be taken into account in research examining the relationships between geological hazards and health outcomes. A range of methodological approaches are discussed and evaluated which may allow these complexities to be addressed in future research.

2. Discussion

2.1. Characterising heterogeneity of geological variables

The aim of a geological survey is to map variability across a certain domain (sample area), providing a distribution of a variable or variables
(i.e., concentration of metals in soil) in space and time. Essentially, a robust sample plan for the survey will reflect the purpose of the investigation, for example whether the map is to make local predictions across the domain, detect the presence/absence of certain components within the domain, or monitor whether the situation has changed over time (and space). When considering the vast number of exposure scenarios possible in the environment, within different environmental domains (e.g., air, soil/food and water) and via assorted routes (ingestion, absorption and inhalation), a broader perspective may need to be employed to identify the characteristics of the study area.

The traditional approach to map soil within a domain is to conduct a survey and collect soil samples for analysis, either in the field or in the laboratory, but sampling strategies are often defined by practical limitations such as funding constraints or logistical impracticalities. Geostatistical modelling methods (with or without the use of covariates), such as Kriging (a method for spatial interpolation), can be applied to investigate spatial variation in observations across the domain of interest, and importantly, to make use of this variation (spatial autocorrelation) to provide accurate spatial predictions at un-sampled locations. The distribution of soils will be determined by various environmental (e.g., parent rock type, climate, hydrology etc.) and anthropogenic (e.g., farming activities, pollution sources etc.) factors occurring at different spatial and temporal scales. In terms of spatial variation, targeted sampling is often compulsory due to the high cost of sample collection and analysis. If soil in the sampling area is highly variable (heterogeneous), the time needed to sample and costs of analyses will be high in order to obtain a sufficient spatial resolution to capture the variability (Vitharana et al., 2009).

When considering the contribution of certain environmental components within a health-related investigation, it is also crucial to incorporate temporal variation within the domain in order to more accurately estimate the exposure. In studies monitoring air pollution, for example particulate matter within a certain size range (e.g., PM_{10} or PM_{2.5}), data for the particulate burden may be collected at point sources in the study area. This data can then be interpolated using other acquired variables (meteorological conditions, urban architecture, and information on the sources of particulate, for example motor vehicle movement) that will impact the density and distribution of the particulate matter over time and space. This information can be used to create maps, defining the variability of the hazard over the sample area. When used in conjunction with public health policy and exposure limits these outputs can be effective in identifying ‘at risk’ areas where the hazard is greatest.

2.2. Characterising heterogeneity of health outcomes

Epidemiological data can be either primary data (generated for the specific research purpose for which they are being used) or secondary data (generated for a purpose different from that for which they are being used, e.g., routine surveillance systems, or previous epidemiological studies) (Olsen, 2008; Woodward, 2013). The underlying population distribution and, therefore, the distribution of health outcomes are both inherently spatially heterogeneous, as are potential geological hazards. When considering the health impacts of geological exposures, it is clearly important to consider this spatial heterogeneity; thus, epidemiological data should have spatial attributes (Pfeiffer et al., 2008; Rothman et al., 2008).

Routine surveillance data will often include information on the administrative area in which individuals reside, allowing the aggregation of cases to specific administrative areas and the presentation of maps of case counts, or in combination with population data (e.g. from census data), prevalence or incidence (Beale et al., 2008; Lawson, 2006). The use of cross-sectional or cohort studies, in which health outcomes are assessed in individuals (rather than aggregates of individuals), gives greater opportunity to attach precise geographical locations, as geographical coordinates can be recorded for individuals’ homes or alternative locations (Pfeiffer et al., 2008) and hence constrain exposure over time.

2.3. Linking geological hazards and health outcomes

The detection of unexpected health outcomes (often signified by unusually high incidence) in a population, suspected to be caused by exposure to a naturally occurring hazard, may instigate a geo-epidemiological study. Thus, acquisition of epidemiological data will typically be the initial response, followed by the collection of geological information to complement this dataset. The domain of interest needs to be considered from the outset as there is little point in assessing health outcomes in an area where the putative geological character does not vary. Thus, the study area should aim to encompass a range of values for the variables that can be measured to determine the hypothetical hazard. In addition, fundamental issues to consider are the potential mechanism of exposure (e.g. the environmental media in which the hazard exists and the route of exposure) and how the individual’s exposure may vary within the population (e.g. genetic propensity, age, behaviour), both of which can be used to develop a dose–response relationship for the hazard.

To establish correlational relationships between the potential geological hazard and health outcomes, the two data sources (the epidemiological and geological surveys) need to be linked to allow statistical analysis. There are different ways of doing this. Where aggregated health outcome data are available within administrative units, data will be linked at the population level as in an ecological study (Woodward, 2013). This approach requires the environmental component(s) thought to be contributing to the disease to be collectively characterised within administrative areas, for example by calculating mean values for each area. Examining correlations in this way can be less demanding than for individual level studies (Nielsen and Jensen, 2005). However, within administrative units (often defined by political boundaries) the components within the environment contributing to the disease are likely to be highly variable and correlations detected at population level may not exist at individual level. Thus, these studies are useful for hypothesis generation for further study and can provide a useful means for the initial assessment of potential causative agents, but are prone to bias and the “ecological fallacy” (Morgenstern, 1982).

Epidemiological investigations at the individual level provide more detailed evidence of the correlations between environmental exposure and health outcomes, although the acquisition of suitable data is typically more time consuming and costly. Survey methods can be used to collect epidemiological data on health outcomes and exposures in individuals (e.g. case-control, cohort or similar study), but assigning quantitative measures of exposure to the environmental component to individuals is difficult. Ecological exposure data (e.g. mean values within an individual’s area of residence) can be linked to individual level health outcome data, although this may not adequately capture heterogeneity in the environmental component, or individual level exposures (Hatch and Thomas, 1993; Nielsen and Jensen, 2005). Estimating exposure to the environmental component for each individual (e.g. at their home) allows us to directly link exposure and outcome information at an individual level, but is more challenging logistically and incurs greater financial costs (Hatch and Thomas, 1993). In addition, individual exposure estimates may be based on subjective information (e.g. questionnaire responses), with the potential to introduce measurement bias. Where it is not possible to take a physical measurement of hazard exposure for each individual included in the study, geostatistical methods may be beneficial. Geostatistical model-based predictions, such as Kriging, can be used to produce spatially continuous estimates of a value of interest (e.g. concentrations of the environmental component associated with the disease) based on an even coverage of data from the sample area: the spatially continuous estimates can then be used to provide exposure estimates for individuals based on their spatial locations (Goovaerts, 2014).
2.4. Consideration of the complexities in geo-epidemiological studies

The nature of the pathological process involved in many medical geology examples, along with the inherent heterogeneity in geological profiles introduce further complexities for the interpretation of correlations between environmental exposure and health outcomes. Several such complexities are detailed below, along with methodological suggestions that may be used to address them, drawing experience from a range of allied fields of research.

Frequently, long-term exposure to an environmental hazard is required to illicit pathogenic changes, giving rise to a lengthy latent period between initial exposure and disease onset. For example, podoconiosis (non-lymphatic elephantiasis), a chronic, disabling condition which produces swelling of the feet and lower limbs, develops following long-term (generally at least 10 years), bare-foot exposure to particular soil types (Davey and Newport, 2007). Similarly, malignant mesothelioma, which is associated with exposure to asbestos, normally develops at least 15 years after first exposure (Brodkin et al., 2006; Lanphear and Buncher, 1992). Thus, any measurement made to elucidate exposure to the environmental component during investigations must be compared to the likely timing of clinically relevant exposures. More specifically, recording environmental hazards based on an individual’s present residence location may not represent epidemiologically important environmental exposures due to short- or long-distance migrations during their life course, or changes in behaviours and exposure patterns over time (Brodkin et al., 2006). To gain an accurate picture of long-term exposure patterns, qualitative and quantitative assessment methods can be used to gain retrospective insight into individual movements over time. For example, the administration of well-constructed questionnaires, use of in-depth interviews or application of retrospective calendars can provide information on residential behaviours or migrations over time, including spatial and temporal information, with varying levels of detail (Carling, 2012). This can provide important supplementary information to ensure that exposure measurements assigned to individuals are appropriate, although the accuracy of information gained via such retrospective methods can be variable and recall bias is a recognised issue (Carling, 2012; Coughlin, 1990). In ecological studies it is more difficult to address such issues as there is no possibility of including information on individuals’ movements.

Secondly, exposure to the geological hazard may vary over time, in some cases due to a deliberate attempt to avoid the hazard. As an example, chronic exposure to arsenic can lead to skin lesions and skin cancer (Naujoks et al., 2013), but in areas with high arsenic concentrations in water supplies, water can be treated within households, which acts to modify exposure levels and subsequent risk of disease over time (Jiang et al., 2012). These individual or household-level variations in exposure have the potential to weaken observed correlations between environmental components and health outcomes, resulting in incorrect interpretations. To ensure a robust assessment of the correlations between exposure and health outcomes, potential mediating behaviours should be accounted for. Detailed questionnaires should be administered to obtain information regarding relevant behaviours, including information on modified behaviours over time that will impact exposure, with subsequent statistical analyses accounting for these confounders where possible. Again, it is more difficult to address this issue in an ecological study, unless reliable data on the prevalence of mediating behaviours across different geographical areas is available.

Underlying geological profiles are typically also inherently heterogeneous and, thus, the potential for exposure to the geological hazard may vary over relatively short distances. For example, exposure to elevated radon concentrations over long periods of time in occupational settings (e.g. mines) and residential buildings can increase the risk of lung cancer (Darby et al., 2005; National Research Council, 1999; World Health Organization, 2009). Radon gas is produced by the decay of naturally occurring uranium deposits in the soil, and is known to be highly variable over distances as short as 10 m (Badr et al., 1993; Oliver and Badr, 1995). This can seriously complicate assessment of exposure-outcome relationships. In ecological studies, spatial heterogeneity in environmental components within the geographical units used for aggregation (e.g. administrative areas) is not considered, giving rise to the ecological fallacy, as discussed above (Morgenstern, 1982). In individual level studies, spatial heterogeneity over small distances necessitates a detailed consideration of the most appropriate location to assign an exposure variable to an individual. Short-distance variation, in combination with patterns of human mobility, means that individuals are likely to be exposed to different concentrations of the potential geological hazard across time (including long term changes due to migration to new areas, as discussed above, or short term changes due to daily mobility patterns). Daily and weekly activity (short term) patterns will influence the time spent in different areas, where individuals may be exposed to varying levels of the geological hazard. Where a single exposure measurement will be taken for each individual, consideration of the individual’s usual movement patterns (“activity space”) could provide useful information with regards to how long they spend in different areas. Again, questionnaires or interviews can be used to gain this information, or movement diaries can be kept by participants to record activity patterns over a defined time period (Belli et al., 2009). For more accurate spatial information, Global Positioning System (GPS) trackers can provide a quantitative assessment of individual movements over varying time periods (Vazquez-Prokopec et al., 2009). Using current understanding of the disease pathology and epidemiology, these movement data may indicate specific locations where epidemiologically relevant exposures are more likely to occur (e.g. podoconiosis is believed to be an occupationally-associated disease, thus, sampling of the soil could be carried out at the most common occupation related location, such as in the individual’s farming fields). However, even where mobility patterns are considered during sampling, it is important to recognise the spatial scale of environmental heterogeneity; a single sample is not necessarily representative of the overall exposure an individual is subject to.

Where information on the spatial heterogeneity of an environmental component is available, along with a comprehensive understanding of individuals’ movements over time, this strengthens our ability to detect clinically relevant correlations between environmental components and health outcomes. However, mechanisms for exposure vary for different substances, and physical presence in an area where the geological component is present (or present at a high concentration) does not necessarily equate to transfer of the potential hazard to the individual. Personal sampling is common where there is potential for exposure to, for example, particulate matter in occupational settings. In this case the monitoring device (commonly comprised of a collection stage, a filter and pumping mechanism) is fitted to the person appropriately for example near the ‘breathing zone’. The device is tasked to collect the particulate sample according to the appropriate metric (e.g., mass/number/surface area concentration within a size range) to best represent the closest approximation to actual exposures (Donaldson et al., 2010). This information on personal exposure can be linked with other data sources, such as questionnaire responses, to better understand the influence of behavioural factors on actual (quantified) exposure. However, such exposure measurement is not available for all potential geological hazards.

In some examples, the exact component or components within the environment that trigger the disease have not been identified. The geographical distribution of podoconiosis is correlated with the presence of red soils of volcanic origin, high altitude (1000 m or higher) and large seasonal rainfall volumes (above 1000 mm per year) (Davey et al., 2007). Research has suggested correlations between several different soil components and the occurrence of podoconiosis, including silicon, aluminium (Price and Henderson, 1978), zirconium (Frommel et al., 1993), smectite, mica and quartz (Molla et al., 2014), although as yet the specific soil component(s) triggering the onset of podoconiosis are not known. Further information in this case is needed to establish a dose–response relationship, although other factors, such as behaviour,
may alter exposure levels over time, thus complicating the correlation between disease and the triggering component.

Podoconiosis is not alone, the exact mechanism of disease initiation is unknown in many cases, which raises the issue of the bioavailability of environmental components within the human body. These points highlight the importance in differentiating between correlation and causation. In cases like podoconiosis, where broad correlations have been made between the presence of the disease and soils derived from basaltic (volcanic) deposits, no one ‘trigger’ component has yet been identified. In these cases, a greater understanding is needed of all confounding factors related to the disease (including genetic susceptibility) in order to build a robust model of the disease aetiology. Linked to this, bioavailability studies are needed to illustrate plausibility that the geological components found to be correlated with the disease could be the causative agent, as opposed to a confounder or spurious finding. In another example, volcanic ash from Montserrat in the West Indies is known to contain crystalline silica (Baxter et al., 1999), a human carcinogen, but its bioactivity in the human lung is thought to be moderated by other compounds in the environment (such as iron). Standard in vitro tests employed to determine the relative toxicity of samples of volcanic ash in the lung have not found the same adverse reaction as pure phase crystalline silica (e.g. Wilson et al., 2000). These in vitro toxicity tests do not necessarily identify the potential for development of disease over long time periods. Understanding the reactions that occur when mineral particles enter and reside within the human body is essential to unravelling disease pathogenesis, even when cause and effect have been identified. Whilst they do not negate the need to monitor the hazard within the environment, methods of enhancing our understanding of both the presence and progression of disease include: for example, analysing biological samples (urine, blood, hair, nails, exhaled breath etc.) for biomarkers and more advanced toxicological studies to elucidate mechanisms of disease.

3. Conclusions

The complex and lengthy exposure-response pathways involved, small-distance spatial heterogeneity in environmental components and a range of other issues mean that interdisciplinary approaches to medical geology studies are necessary to provide robust evidence. Geological and epidemiological methods must be linked and the spatial component always be addressed. These methods may be supplemented with quantitative and qualitative approaches, such as questionnaires, diary or calendar based approaches, or GPS tracking to capture spatial and temporal variations in exposure due to movement patterns or migration. Taking an individual level approach is the most appropriate to enhance our understanding of both the presence and progression of disease include, for example, analysing biological samples (urine, blood, hair, nails, exhaled breath etc.) for biomarkers and more advanced toxicological studies to elucidate mechanisms of disease.

Competing financial interest declaration

The authors declare no competing financial interests.

Acknowledgements

Many thanks to Professor Gail Davey for providing input on a draft of the manuscript. N.A. Wardrop was funded by the Medical Research Council, UK (project MR/J012343/1).

References


References


