What Is the Exposome and Why Does It Matter to Your Health?

Professor Ian Mudway

27th October 2022

Introduction

In today's presentation I want to introduce you to the concept of the exposome and its related discipline of exposomics. However, before delving into dictionary definitions, I'd like to assure you that we will be dealing with a weighty topic which encompasses both life and death. This means that we can frame the topic between two absolute certainties which is always a good way to start. Everyone here today or listening online has been born and eventually, irrespective of the wonders of modern medicine, each and every one of us will one day die.

That span of time from womb to tomb is our focus. Specifically, how we age, develop disease, decline, and ultimately die and how our environmental exposures in their very broadest sense, contribute to these processes across our entire life. Framed in this way this presentation may seem a little bleak, but I'd encourage you to memento mori (remember that you [have to] die) so that we may reflect on the best way to live. This is something that we're not great at in Western cultures. It just requires little mental gymnastics to view this in a more positive light, as a new emerging science that aims to ensure that our environments promote optimal and equitable health and well-being across the population. Less obsessed with identifying every environmental hazard and risk that can cause us harm (because they are numerous, as you will see), more a science to deliver improved public health. This allows us to live longer, healthier lives and reduces the unnecessary burden of chronic disease to both the individual and the state.

But definitions can be helpful, so let's begin. The exposome is defined, according to the 6th edition of the dictionary of epidemiology as *"a potential measure of the effects of life course exposures on health. It comprises the totality of exposures to which an individual is subjected from conception to death including environmental agents, socioeconomic conditions, lifestyle, diet and endogenous process."*

Let's pause to unpack this a little.

Life Course Exposures

Every time we breathe, eat, drink, exercise, go to work, socialise, shower and sleep we expose ourselves to potentially harmful substances from air pollution to synthetic chemicals, pharmaceuticals, UV, the food we eat, the alcohol we drink, excessive noise and the microorganisms we encounter throughout our daily lives. These exposures are obviously highly variable, not just as we progress from childhood to adulthood, but quite frankly between morning, evening and night of every day. Can this kind of information even be captured in a meaningful way across the span of a human life and when are the relevant exposures – in utero? In early life as organ systems mature? During late middle age when our defence and repair mechanisms degrade? Can we determine the sum of exposures across this whole period and what should they be linked to? It would already be too late to understand their contribution to disease development if the disease was already fully developed. You would really need to be showing associations with the early hallmarks of disease aetiology and progression, prior to the development of overt disease. For example, reflecting on cardiovascular disease evidence that exposure was linked to increased cholesterol, arterial plaque formation, inflammation, hypertension etc., - intermediate biomarkers. So, when we're talking about the exposome we're moving beyond blunt measures of health and timing is everything

Then we move onto the specific domains of exposure**: environmental agents, socioeconomic conditions, lifestyle, diet, and endogenous process.** Clearly there's some overlap here. We know for example that

being poor is a significant risk factor for premature death, for additional years of poor health with a wide range of health conditions. Poverty itself represents an amalgam of environmental factors, from poor housing, poor diets, increased exposures to pollution, increased psychological stress and so forth. So, is it possible to precisely dissect out the contribution of a particular environmental stressors from another when they are so intrinsically linked? Then we have endogenous processes; internal biological mechanisms that modify and reflect the nature of exposures, through inflammation, metabolism, detoxification, the action of gut microflora etc. An additional level of complication.

I make these points now, not to derail the whole enterprise of understanding the environmental contribution to chronic disease, but simply to highlight that this is no minor undertaking and that capturing the human exposome and understanding it's interaction with our underlying genetics in disease development will be one of the great challenges of 21st century science. For now, it would probably be useful to reflect on the background to the development of the concept

Pre- August 2005 there was no concept of the exposome, at least not in the way I'm discussing in this lecture, as a comprehensive set of exposures over an extended period. It's not that people didn't appreciate that chemical hazards in the environment caused disease. The occupational literature was very mature, chemical vapours, metal fumes, toxic fibres, fumes and particles - Theophrastus von Hohenheim (a.k.a Paracelsus, father of toxicology) published on the diseases of miners in 1533. Certain environmental stressors such as air pollution were known to be associated with both long and short-term health effects and premature death. Also, the interaction between chemical carcinogens and cancers were well established, with established biomarkers of dose – DNA adducts, within tumours, but throughout the 1990, into the 2000's this was not the focus of attention. The late 1990s to the early 2000s represented the genomic era. The Human Genome Project was launch in the United States in 1990, with the first draft of the human genome sequence completed in 2000, published in Science and Nature in 2001 and the completed sequence was finished in 2003 (all 3 billion base pairs). The next phase was to examine the variation in the human genome using population level sequencing to allow the examination of the relationship between genotype (single nucleotide polymorphisms or SNPS, variant forms of a gene or alleles) and phenotype (an individual's observable traits) in genome-wide association studies. However, whilst certain traits or diseases could be largely explained from genomic variation, such as eye colour and type I diabetes, this was not the case for many of the major non communicable diseases such as cancer, cardiovascular and pulmonary diseases and even longevity. Genetic variation only explained a minor proportion of the risk observed, suggesting that the environment played a dominant role. Why does this matter?

Noncommunicable diseases (NCDs) are the leading causes of mortality and morbidity globally. Approximately 30% of NCD-related deaths are considered premature, occurring before the age of 70 years, and approximately 85% of premature deaths occur in low- and middle-income countries. The World Health Organization (WHO) has identified 4 main types of NCDs that contribute the greatest burden. They are cardiovascular diseases (CVDs), cancers, chronic respiratory diseases, and type II diabetes. In 2016, NCDs accounted for 41 million of deaths worldwide, with 44% of these deaths attributable to CVD, 22% to cancer, 9% to chronic respiratory disease, and 4% to diabetes. For all of these conditions underlying genetics represents only a fraction of the observed risk. Thus, our environmental exposures appear to be overwhelmingly more influential in determining our mortality risk. But exposure to what? That's the critical question. Though we hardly start with a blank slate, as we already have a firm understanding of the role played by unhealthy diets such as cigarette smoking, excessive alcohol intake etc. However, this only represents a fraction of our complex environmental exposures and addressing these is therefore critical to understanding NCD risk and improving global health.

Deaths and ill heath due to NCDs increase as the population ages which is important as global life expectancy will continually increase throughout the current century, with ever increasing numbers of vulnerable individuals such as children, adults of reproductive age and the aged themselves. But these increased risks are not shared equally, and the poor are disproportionately affected.

Consider for example the current situation in England, using data from the ONS for 2017-2019. If you compare individuals living in the 10% most deprived areas with those in the 10 most affluent regions, the current difference in life expectancy is 9.4 years for men and 7.7 years for women. This difference becomes even more stark when you examine the years of healthy, or disability free life, with the difference being 19.3 years for women and 18.6 years for men. I can't think of a much clearer example of the impact of our environmental exposures on our health.

Underlying these differences are profound increased risks of NCD deaths, from cancer, cardiovascular

disease, dementia, and respiratory disease with the latter being a more than twofold difference between those in the most deprived areas in comparison to the least deprived ones.

So, we're undoubtably in the grip of a global epidemic of NCDS which has resulted in significant human costs. However, it is also crucial to reflect on the economic costs of failing to tackle this challenge – which in a sense will require us to deliver better diets, lifestyles and environmental exposures. The cost of continued underinvestment in the fight against NCDs has been estimated at a US$47 trillion loss in gross domestic product globally from 2011 to 2025.

Parallel to the increase in the global ageing population, we have seen profound changes in population distribution over the last century that has contributed to substantial changes in population level exposures. Perhaps the most significant of these has been the global movement of population from rural areas to the city, especially to the emerging megacities within low- and middle-income countries. This is illustrated for example by the up map showing global population densities, with each skyscraper represented a city. The UN estimates that it was in 2007 that the number of people living in urban areas first overtook the number in rural setting and by 2050 its estimated that 2/3rd of the global population will live in urban areas: approximately 7 billion people.

This rural to urban shift has occurred during a period of unprecedented expansion in chemical synthesis which again has increased the complexity of external chemical exposure and the numbers are quite arresting. Since the chemical revolution of the 1800s over 200 million chemicals have been reported in the literature, with the currently global inventory of chemicals licenced for manufacture and sale being 359,206. Only around 7% of these have been registered with REACH, the Registration, Evaluation, Authorisation and Restriction of Chemicals, the body which evaluates potential hazards and risks to the population. This itself does not reflect the totality of our chemosphere, which includes chemicals excreted as metabolites, and/or those that undergo biotransformation by microbiota within the environment. The chemicals that we are exposed to has therefore expanded alarmingly over a relatively short period of human history and this includes pesticides, plasticisers, phthalates, heavy metals, manufactured nanoparticles, microplastics, pharmaceuticals etc.

It was against this background that Christopher Wild published his seminal editorial in 2005 in Cancer Epidemiology Biomarkers and Prevention. This represented a call to arms for a greater focus on the role of environmental exposures on disease development, to complement the work examining the role of genotype on the risk of chronic disease. At the time, several key components that were necessary to advance work in this area were beginning to coalesce: (1) the development of high throughput techniques to evaluate genetic variation in large numbers of subjects, (2) new mega cohort resources (groups of individuals recruited into long-term research projects) such as the UK Biobank where genotype data would be available. This could be linked to health data over time, and used with the collection of bio samples (blood and urine) to evaluate biomarkers of exposure and adverse biological responses, (3) advanced technologies for looking at global biological responses in gene regulation (the 21,000+ genes in the human genome, transcriptomics), protein responses (the 18,407 proteins (HPP), proteomics) and changes in the end metabolic products of these responses (217 920 compounds (Human Metabolome Database), metabolomics); and (4) finally reflecting the complexity and the scale of such gene-environment interaction studies, this required 10-100,000's of individuals who used advanced computing, statistical and bioinformatic approaches. This absolutely qualifies as big data. Having reread the article in putting together this presentation it is remarkable to see how current and prescient the thinking was and yet the paper was not marked by fanfare, but almost by a deafening silence for five years after publication. It's not that the importance of gene-environment interactions were not appreciated, but it was the era of the big 'G' and small 'e' and the paper somewhat sank from view.

Momentum around the exposome concept picked up around 2010 with the publication of an article in Science by Steve Rappaport and Martyn Smith, from the University of California, Berkley, which focused on how to implement the original vision outlined by Wild. Here the authors proposed the adoption of two complementary strategies: a *"bottom-up"* and *"top-down"* approach. In the former method all chemicals in each external source of a subject's exposome are measured at each relevant time point in air, water, or diet etc. This represents an enormous undertaking with measures of the complexity of the exogenous chemical environment that can be related to a health endpoint or a range of endpoints which range from a biological response, a symptom, or the presence of an overt disease. This approach would miss critical aspects of the individual internal chemical environment such a measure of phycological stress, metabolism, inflammation etc, and would not be sensitive to modifying factors related to age, sex and ethnicity. The alternative "top-down" strategy, which might be framed as an agnostic biomarker approach, effectively uses the individual themselves as the environmental sensor and employs comprehensive chemical profiling of an individual's

blood, or urine, or tissue sample if this was available at relevant time points. This would produce an array of potential correlates which would be focused on known toxicant classes such as metals/metalloids, reactive electrophiles, endocrine (hormone) disruptors, immune modulators, agents that bind to cellular receptors etc. However, these also frame their presence within the context of the induced biological response which permit exposome-wide association studies, and which provide an array of individual chemical entities or groupings of chemicals that relate directly to a person's internal chemical environment and health status. Once important exposures have been identified in blood samples, additional testing could determine their sources within the external environment and methods could be adopted to reduce them. This may seem overly ambitious by US National Institute of Environmental Health Sciences (NIEHS) and the UK's own National Phenome Centre, for example, have developed methods for identifying thousands of compounds in blood, urine, saliva, water and household dust. Thus, the technology has now evolved to a point, in parallel with the multiple 'omic technologies to evaluate biological responses to allow such ambitious undertakings.

These ideas are summarised in the following figure. The Bottom-up approach focused on measures of the external environment, both general and specific through direct physical measurement, surveys or enhanced modelling approaches, which can be linked to health endpoints, or intermediate biomarkers on the pathway to disease development. In contrast, the top-down approach focuses on changes in the internal environment assessed through multi-omic technologies such as transcriptomics, proteomics, changes in the microbiome, the metabolome and the epigenome (chemical modifications to our DNA that effect gene regulation), allied to established biomarkers of chemical exposures and response which is again linked to established disease, biomarkers, and increasingly organ specific, or whole-body imaging. Whilst these global profiling approaches are often conducted one at a time, there is increasingly a move toward integration of the data across these various platforms by clustering data together with information on exposures and disease. This is undertaken to drill down further into causal pathways and toxicants to adverse outcome relationships.

Whether such investigations start from disease in search of a cause (top-down), or it starts from the other end (bottom-up), there is always the need to connect the external exposure with a disease outcome through a set of intermediate biomarkers. The simplest approach to do this is the "meet-in-the-middle" (MITM) approach. This consists of measuring intermediate biomarkers (often with an agnostic 'omic investigation, but sometimes informed by prior toxicological knowledge) and relating them retrospectively to measurements of external exposure and prospectively to a health outcome (disease, or aging, or other outcomes) in the context of a longitudinal investigation. Simply put, this is administered to demonstrate adverse pathways that are induced ahead of disease presentation.

Often, these approaches are fully agnostic. By this I mean that there is no prior hypothesis, and the studies are effectively purely observational. This is often described as hypothesis generating, but this does not have to be the case and in fact from a personal point of view I think it's a mistake not to acknowledge our extensive prior knowledge from decades of toxicological and disease-focused mechanistic science. We can therefore build frameworks that attempt to explain the disease pathway from toxicant to adverse outcome, and then test these or use them to add coherence to the results arising from large scale exposomic studies. For example, if we consider cancer, we know what the key characteristics of a carcinogen. They are: (1) electrophile either directly or after metabolic activation; 2) genotoxic; 3) they alter DNA repair or cause genomic instability; 4) induce epigenetic alterations etc. We also know the condition necessary to promote cancer development which is the Hallmarks of Cancer. This is defined in Hanahan in 2011 as the induction of pathways that resist cell death, sustain cell proliferative and replicative immortality, evasion of growth suppressors, activation of invasion and metastasis, and angiogenesis. These therefore provide a clear trigger, or molecular, which initiates event-defined stages (key events) in disease progression. In this case this results in a cancer as an adverse outcome, but the same approach can be applied to other NCDs.

Just over a year ago Annette Peters, Tim Nawrot and Andrea Baccarelli published a paper that builds upon earlier papers in the journal Cell which focused on the biological mechanisms underpinning cancer development and ageing in an attempt to reflect on how environmental insults cause or contribute to disease. They identified 8 hallmarks of environmental insults that can be adopted in the development of adverse outcome pathway approaches: 1. Oxidative stress and inflammation, 2. Genomic alterations and mutations, 3. Epigenetic alterations, 4. Mitochondrial dysfunction, 5. Endocrine disruption, 6. Altered cell communication, 7. Altered microbiome communities, 8. Impaired nervous system function. What was notable about this was the overlap with the hallmarks of ageing- genomic alterations and mutations, epigenetic alterations, mitochondrial dysfunction and altered intracellular communication. These hallmarks of ageing, along with oxidative stress and inflammation, are known to increase as we grow older. Given that NCDs are diseases of age this clearly raised the possibility that a whole range of environmental stressors are

contributing to an accelerated ageing phenotype. Indeed, a number of biological clocks to assess biologic versus chronologic age, based on gene methylation, or changes in metabolic profiles as we age are now available to address this question. But at its heart is the contention that an individual, say a 53-year-old slightly jaded biochemist, who has lived in Central London for the last 25-years may be walking around with a 60-year-old cardiovascular system due to the access wear and tear of my urban lifestyle.

These hallmarks of environmental effect can then be mapped to different organ systems in the body to highlight the type of intermediate biomarkers that will be most relevant to organ-specific adverse pathways contributing to disease. In lungs for example, the induction of inflammation, and changes in the microbiome, perhaps necessitates the addition of oxidative stress which would result in brain stress hormone release, neuroinflammation etc.

There are challenges in attempting to integrate adverse outcome pathways into the exposome concept and to illustrate this I thought I'd focus on the pollutant gas NO_2. This is mostly because many of you will have breathed unhealthy concentrations of this gas simply getting here this evening. The first issue is one of temporality and sensitivity. If you have a pre-existing disease, be that respiratory, such as asthma, or COPD, or cardiovascular, you are likely to be more vulnerable to acute adverse responses to this gas. It may trigger, shortness of breath cough etc. This is because NO_2 is a strong oxidizing and nitrating agent, and we have a clear molecular initiating event through the imposition of oxidative stress which is associated with this inflammation. However, with all toxicants that we are exposed to, we have to consider inter individual variations in how the gas is absorbed into the body, how it is distributed following its uptake into the lung, the subsequent metabolism and in the case of compounds that undergo xenobiotic metabolism, its exception. These processes will vary based on our underlying genetics and pre-existing disease.

But the challenge comes when we try to look at the longer-term effects of recurrent challenges. Whilst the initial trigger may be the same, the tissue response is likely to trigger particle through adaptation to the persistent irritation. Soon one might see a transition from acute to chronic inflammation and remodelling of the airway epithelium alterations in lung function etc., many of which are part of normal ageing within the lung.

In truth of course, things are not so simple. We don't breathe one pollutant, but a complex cocktail that varies in concentration throughout the day including both primary emissions and secondary chemical species formed through atmospheric chemistry.

So, in a sense what the lung sees is multiple agents, potentially with different molecular initiating events, different modes of action combining to produce both acute and chronic effects. Thankfully however, there are commonalities such as oxidative stress, inflammation and cell injury. .I can illustrate this in the next study which adopted the meet in the middle exposomics approach to link adverse outcome pathways related to air pollution to downstream cardiovascular risk.

This study made use of the EPIC-Italy study where DNA and blood samples were available 17 years prior to the diagnosis of cardiovascular disease. Proteomic analysis demonstrated that prior to disease air pollution exposures were related to IL-17 expression, the DNA methylation analysis demonstrated increased methylation within genes related to oxidative stress; the generation of reactive oxygen species and glutathione metabolism. As air pollution within the cohort was already known to be associated with CHD and CHD has been associated with systemic inflammation and oxidative stress the authors were able to infer that induction of the former AOPs were related to the subsequent disease development. These studies are based on exposure estimates from models which are generally based on annual exposures at a point location and there has been some concern about how accurately these estimates capture an individual's true exposure.

One way of getting individualised estimates of an individual exposure is via personal monitoring. Here I'm summarising a study by my colleague Ben Barratt in which patients with COPD were given mobile personal pollutant monitors to carry over extended periods to relate to their symptoms which included periods of exacerbation. The portable air pollution monitors from Rob Jones's group at Cambridge University were validated against gold standard instruments prior to use and representative data is shown in the next slide. This includes noise, temperature, a range of pollutant metrics and below parallel information on symptoms recorded by questionnaire or peak flow testing. The results, published in the European Respiratory Journal in 2021 clearly showed increased risk of exacerbations with NO2 exposures illustrating that exposures were also related to perceptions of breathlessness, cough, and sputum production.

For larger scale investigations modelling is typically employed to estimate exposures. Models for Europe and the UK are available for annual exposures to a range of pollutants, typically PM2.5 and PM10, as in the left-

hand panel and NO2 for the UK. These models are slightly different, based on different approaches which either relate measurements to land use to derive statistical models to predict exposures, or use information of emissions allied to atmospheric dispersion and chemistry. However, in both cases the models allow you to link an exposure at a given point, be it a residential address, a post code, or as an estimate of long-term exposures so that relationships to health endpoints can be derived. These models are usually related to criterion or legally regulated pollutants but there is much more work to be done in this area.

Clearly using static models of exposure is at best a crude estimate, because the population don't spend their days parked on their front doorstep. Clearly you can overcome this with personal monitoring, but the challenge will always be scale. Personal monitoring isn't currently, and I suspect will never be feasible for the whole population, so the alternative is to enhance our current modelling approaches to reflect the mobility profiles of the population. There are several ways of doing this, but again they remain imperfect refinements. We can learn more about how the population moves around our cities and then link this with questionnaire information to estimate exposure. For example, the following map is based on TfL's travel demand survey. We can see from it the flux of people entering and leaving the city on a given working day and this means that we can estimate where they are, how long they are in transit, how long they spend at work and at home. With this information we can begin to build more comprehensive measures of exposure, but this clearly requires knowledge of exposure likely to occur on the London Tube, whilst sitting in traffic, and finally whilst at home where the nature of exposures will be significantly different from the outside environment. These are not insignificant challenges, but significant progress has been made in recent year and importantly the predictions of exposures in mobile populations can be validated in smaller scale personal exposure studies.

The situation in air though is clearly very complex, because not only do we have mixtures of primary pollutants derived from the same source and therefore highly correlated, but also that these exist within an even more complex mixture of secondary inorganic and organic component which are formed through atmospheric transformation processes. So, the question is, how do you show whether one component is worse than another? This has very significant policy ramifications.

Can exposomics approaches help here? The following study examined the impact of air pollution exposures in individuals (healthy individuals and subjects with COPD and IHD) exposed separately on two occasions in London, once in a high diesel exhaust environment – Oxford Street and the other a lower pollution site walking around the Round Pond in Hyde Park. All subjects had personal measurements of their air pollution exposures made which reduces the very high correlation between certain components and they provided blood samples at various points during their exposure day. In this study the authors use sequencing to examine circulating microRNAs in the blood post HP and OX street, which have been proposed to act as useful biomarkers of extra pulmonary effects. Several were identified that showed differential expression between the two exposures, some of which have been formerly notified as being indicative of adverse effects on peripheral organs. The key observation is presented in the following Venn diagram where the relationship between individually measured components of the air pollution aerosol – PM10, black carbon, NO2 etc, which are often highly correlated in air with specific microRNAs are shown. What is evident is that there is very little overlap, suggesting that biologically at least these different pollutants are acting via different pathways.

There are several clear challenges for the integration of the exposome concept into policy. First, there clearly needs to be a significant refinement in our assessment of exposures to environmental toxicants which should include enhanced work on personal monitoring and investment in establishing good internal biomarkers of exposure. This will significantly help to deal with issues of exposure misclassification, which is just a polite way of saying getting exposures a little bit wrong. We need to think much more about how we deal with complex mixtures of toxicants. That may be highly correlated in the environment but may act through entirely separate biological pathways in the body – again a very strong argument for using the individual as the environmental sensor. Ultimately, we need to move beyond simply identifying adverse toxicants – we need outcome pairing to establishing a dose response relationship within the population that can be used for health impact and economic assessments. We need to exploit the meet in the middle methodology to be able to identify earlier biomarkers of progression to chronic disease states and finally, as always, we need much tighter standardisation of methods and validation of key findings. Finally, what would really help would be better biomarkers of long-term exposure. Currently looking at circulating adducts between proteins and chemical metabolites (adductomics) provides us with maybe a monthly estimate, epigenetic promises to provide longer term signatures, as has been demonstrated for heavy metals and cigarette smoking. We need much more work on the role of biological ageing which may itself provide a useful chronic biomarker of adverse effects on the body and finally we maybe need to think about samples from accumulative tissues, lymph nodes, body, hair, teeth etc.

There is some good news in progressing work in this area. Europe has already invested heavily in the construction of large, 10,000 plus cohorts, including the UK Biobank with baseline collection of biological samples, genotype information, linkage to health records and imaging, plus and follow-up of participants. Indeed, earlier this week the public were encouraged to enlist in the new **Our Future Health Study**, which aims to establish a 5-million strong volunteer cohort enabling research intended to improve the early detection of chronic disease, including a strong focus on ethnicities poorly represented in the original Biobank study. The ambition of this project creates a significant resource to support future work in exposomics and to increase our understanding of the environmental contribution to NCDs.

So, in conclusion the exposome has arrived with the ambition to provide almost a route map to understand how our environment impacts on our health. In this presentation I've only been able to scratch the surface of the work beginning in this area and how it will contribute to the shift from the expensive management of pre-existing disease to disease prevention, which will be vital if we are to provide comprehensive health provision to the population moving forward. There are clearly challenges, both methodological (the scale of the challenge, getting a grip on lifetime exposures etc) and in terms of policy development, where the tools to deliver improved public health are via public education and regulation. But before we head to the questions, I'd like you to just reflect on one single fact from the lecture, that in England today (the world's 5th/6th largest economy) there is an almost 20-year gap between to years of healthy life between to wealthiest and the poorest members of our society, reflecting the totality of their experience, lifestyle and lived environments. The exposome really matters to these individuals, even if they have never heard of the term.