The origins of human disease: a short story on “where diseases come from”

Johan P Mackenbach

Most of public health is based on the working hypothesis that disease is caused by exposure to noxious factors in the external environment. While this approach has produced great successes in primary prevention, a general theory of the origins of human disease cannot be found in the textbooks of public health or epidemiology. This paper suggests that, in all its manifestations, disease is a reaction of the human organism to, and/or a failure to cope with, one or more unbalancing changes in its internal environment. These are caused by one or more unfavourable exchanges with the external environment and/or failures in the structural and functional design of the organism. In the final analysis, human disease is attributable to the dependence of organisms on a fundamentally hostile external environment and to unfortunate evolutionary legacies. While this sketch of a theory suggests that there will ultimately be some hard limits to primary prevention, it also helps in identifying possible new approaches to prevention, including interfering with disease mechanisms, and remedying human organisms’ design failures.

In early spring 2004, I had the opportunity of visiting Crete for a project meeting funded by the European Union. We had lunch in a village, and were served a Cretan meal that included one of its more mysterious ingredients: green purslane leaves salad. The fame of the Cretan diet was established in the seven countries study, which showed that men in Crete had incredibly low rates of cardiovascular disease: less than 10% of the rate in Finland and the USA. This has been ascribed to their consumption of olive oil, fruits, vegetables, and red wine, and perhaps wild greens such as purslane that are rich in cardioprotective compounds such as α-linolenic acid and flavonoids.

Stories of how our diet influences our health are part of a wider “ecological” view of what causes disease in living organisms. As many textbooks of epidemiology and public health routinely explain, this tradition goes back to Hippocrates’ treatise “Airs, waters, places”, and has inspired empirical studies of variations in disease rates between different localities until late in the 19th century. Such studies are still important sources of knowledge on determinants of disease, as shown by the seven countries study, and by Doll and Peto’s famous study on the causes of cancer. This used the 10-fold to 100-fold worldwide variation in incidence for most cancers, and the changes in incidence upon migration from one environment to the other, to point to the potential avoidability of this disease. This ecological view has become a central part of the paradigm of public health as it has developed since the 19th century, and has laid the foundation for many successful primary prevention measures.

And even further reductions in disease rates may be feasible: many public health reports have trodden in the steps of Doll and Peto, and argued that enormous reductions of disease rates are possible if we would reduce exposure to a number of current disease determinants. Every four years, the Dutch National Institute for Public Health and the Environment publishes the population attributable risks for a range of risk factors in the Netherlands. The most recent estimates are that 15% of all deaths are attributable to smoking alone, 10% to diet, and another 5% to obesity. Recently it has been suggested that regular consumption of a “poly-meal”, an admittedly improbable combination of healthy ingredients, will reduce cardiovascular morbidity and mortality by more than 75%.

As these examples show, public health is an exciting enterprise, but one wonders where the limits to primary prevention lie. These estimates suggest that after the successes of the first epidemiological transition, in which incidence and mortality from infectious diseases were reduced by more than 90%, we may one day succeed in nearly or totally eliminating their successors, cardiovascular disease and cancer, leading to a second major transition. Is this plausible, and if so, what will happen then? Can we push back disease rates indefinitely by primary prevention? Do all diseases originate in the environment? Is there a real possibility of a counterfactual “paradise”, in which human beings would not be exposed to noxious environmental circumstances, and remain totally free of disease?

These rather naive questions all boil down to an even more simple but equally unanswered question: where do diseases actually come from? Unlike physicists, who spend much of their research budgets in trying to find a “unified field theory”, which should be capable of describing nature’s forces within a single and coherent framework, epidemiologists and public health researchers do not show much interest in finding a unified theory of disease causation. One will look in vain for such a theory, or even a...
brief general introduction on the nature and origins of disease, in the textbooks of public health and epidemiology (and, for that matter, in textbooks of pathology or general medicine). Nevertheless, a few possible ingredients can be sketched—and will be offered here for further discussion and elaboration. In this exercise, we will search our way back from manifestations to onsets of disease, from onsets to direct causes of disease, and from direct causes to origins of disease (see fig 1).

MECHANISMS OF DISEASE: REACTIONS TO UNBALANCING CHANGES IN THE INTERNAL ENVIRONMENT

Medicine has long been dominated by ontological theories of disease—that is, theories in which diseases had an existence independent or separate from the suffering organism. The naive question “where do diseases come from?” actually reflects this popular error. It was only in the 19th century that Rudolf Virchow declared that “diseases have no independent or isolated existence; they are not autonomous organisms, not beings invading a body, nor parasites growing on it; they are only the manifestations of life processes under altered conditions”.4 Since then, disease is seen as an attribute of an organism—a set of pathological manifestations that we can classify in different categories (disease entities), but that should be viewed as reactions of the organism to certain pathological onsets. The link between pathological onsets and clinical manifestations is provided by so called disease mechanisms.15

Box 1 gives important examples of pathological onsets, disease mechanisms, and resulting diseases. Despite the enormous variety in mechanisms, most can be understood as reactions of the organism to, and/or a failure to cope with, some unbalancing change in their “internal environment” (Claude Bernard’s milieu intérieur). This fits within a systems theory of life: living organisms can be seen as systems that are in a state of dynamic equilibrium, not with their external environment but internally. To survive, they must continuously exchange energy and matter with their external environment, while keeping the composition of their milieu intérieur stable within narrowly defined limits.16 17 They adjust to changes in their internal environment through a variety of homeostatic mechanisms, such as behavioural responses, immune reactions, and detoxification procedures. According to this view, disease results when changes in the internal environment temporarily or permanently exceed the capacity for adjustment of the organism, or when a defect in the adjustment mechanism leads to an inadequate response.18

It is beyond the scope of this paper to review even a tiny fraction of the explanatory insights that research into disease mechanisms has generated. In the case of infectious diseases, the “unbalancing change” in the internal environment is invasion by micro-organisms. When the immune system does not succeed in timely elimination of invaders, injury to cells, and organs will occur, both directly (as a result of the invasion) and indirectly (as a result of the actions of the immune system).15 18 The uncontrolled cell growth that defines cancer, results from cyto genetic abnormalities caused by mutations in somatic DNA (attributable to random replication errors or mutagenic factors that inactivate a tumour suppressor gene or activate an oncogene), co-occurring with a failure of the organism’s defence mechanisms (for example, DNA repair, apoptosis of aberrant cells, immune reactions against aberrant cells).15 18 Multistage
The origins of human disease

Theoretical causes of disease are based on the hypothesis that, because of the elaborate mechanisms controlling cell growth, several mutations in a single cell are necessary to produce a cancer.19

DIRECT CAUSES OF DISEASE: UNFAVOURABLE EXCHANGES AND DESIGN FAILURES

Because changes in the internal environment are induced either by exchanges with the external environment, or by an autonomous derailment of internal mechanisms, all diseases must be attributable to either unfavourable exchanges with the external environment, or to failures of the design of organisms, or to a combination of the two.

Exchanges of energy and matter with the external environment are usually advantageous, because they provide the organism with basic necessities (food, oxygen, warmth, and water).20 Problems arise, however, when the intake of these basic necessities is insufficient or excessive, or when the organism encounters or ingests harmful forms of energy and matter (physical, chemical, biological, psychological noxae).21 It is important to note that these exchanges with the external environment are usually active processes, and are therefore determined as much by the behaviour of the organism as by events or circumstances in the environment.

The interaction between organism and environment blurs the distinction between the two: the environment of an organism should perhaps even be defined as the set of external conditions with which it interacts.22 It is worth noting that the active term exchanges with the environment removes the need to separately distinguish behaviour as a cause of disease.

Despite the incompleteness of our knowledge about the aetiology of diseases, it seems reasonable to suspect that most diseases can, and often do, arise as a result of unfavourable exchanges with the external environment. This is immediately obvious for pathological onsets and disease mechanisms that have been defined in terms of the consequences of a specific exchange with the external environment (for example, infectious disease, nutrient deficiencies, mechanical wear and tear, intoxication, injury). But unfavourable exchanges with the external environment probably also play some part in the aetiology of most other disease mechanisms: autoimmune reactions, allergic reactions, metabolic disorders such as diabetes mellitus, atherosclerosis, cancer, ...15 23

While this comes close to public health's ecological paradigm, it is important to note that there is no scientific basis for the claim that all cases of all diseases can be attributed to unfavourable exchanges with the external environment. Some cases of all, and many cases of some diseases may be entirely attributable to an autonomous derailment of internal mechanisms, caused by an apparent failure in the (genetically determined) structural or functional design of the human organism. Hereditary DNA abnormalities are a clear example of diseases where most cases are probably attributable to such a design failure only. Huntington's disease is thought to be attributable to a number of CAG repeats on chromosome 4, the expression of which does not seem to be influenced by any external factors (although the mutation itself could theoretically be because of environmental exposure of an ancestor).15 24 The importance of design failures in the causation of disease goes much further than their role in rare inherited disorders, however. At least three types of commonly occurring design failures can be distinguished.

The first is that human organisms cannot always cope effectively with unfavourable exchanges with the external environment. We dispose of a large variety of adjustment mechanisms, ranging from behavioural responses and protective structures to detoxification mechanisms and immune responses. Some failure of these adjustment mechanisms therefore can usually be identified when disease occurs—and is probably involved in, and to provide the theoretical rationale for, many gene-environment interactions.25 26

But design failures are not limited to those that interfere with our responses to the external environment. Another is that human organisms cannot deal perfectly with entropy.27 Survival requires constant reproduction—of germ cells (for creating progeny) and of somatic cells (for repair and maintenance). During cell reproduction, small random errors occur from time to time, and although we have exquisite repair and search and destroy mechanisms, some errors are not detected in time, and may lead to congenital disease in progeny or to cancer in ourselves.28

And then we have a third design failure: we have limited supplies of vital stocks. It is probable that aging is at least in part determined by cumulative effects of various noxae and our failure to deal effectively with them,29 but not all of aging can be explained in this way. For example, women are born with a limited number of eggs, and when their supplies are exhausted reach menopause. A similar example is that our immune system has only a limited stock of immune cells that can be directed at specific antigens: after we have exhausted these supplies, we cannot cope with new micro-organisms any more.28

There is no need to claim that unfavourable exchanges and design failures entirely determine the occurrence of disease. Random events within cells at the level of molecular interactions, or within organisms at the level of interactions between organ systems or metabolic processes, will partly determine the outcome of the process.22 30 It is therefore an illusion to think that all inter-individual variation in disease occurrence can ever be explained.31

ORIGINS OF DISEASE: HOSTILE ENVIRONMENTS AND EVOLUTIONARY LEGACIES

Why do human organisms have unfavourable exchanges with the external environment? The historical successes of public health show that not all these exchanges are inevitable, but it is unlikely that they can ever be totally eliminated. This is because we are dependent on the external environment, and because the external environment is fundamentally hostile. Living organisms can be characterised as open, non-equilibrium thermodynamic systems that continuously exchange energy and matter to decrease internal entropy.16 17 But energy and matter have to be extracted from an environment that is often hostile, as shown by the fact that the same elements in the abiotic, biotic, and human environments that we are dependent on, can make us sick as well.

Dependence on a hostile abiotic environment takes many forms. The simplest illustration is oxygen: all higher life on
earth is dependent on oxygen, but oxygen is a dangerous compound, as one can see from the fact that it destroys iron. During the processing of nutrients in our bodies, oxygen is released, and this has been shown to contribute to cardiovascular disease and cancer. Fortunately, antioxidants like flavonoids from purslane give a little bit of protection, but this is only partial. All life on earth is dependent on sunlight, but exposure to ultraviolet and other radiation may lead to mutations and cancer. There is also a fundamental hostility, in the form of competition, between different lifeforms. Human beings need plants and animals for survival, but the latter do not always cooperate, and sometimes behave like enemies. For example, plants protect themselves against predation by containing toxins that may make us ill. Bacteria feed on us, and although we have developed sophisticated mechanisms of defence, they are smarter, and continuously succeed in circumventing our countermeasures. And we need other human beings: the prosperity of the Western world that has helped to almost eliminate infectious disease and double life expectancy at birth, is the result of an intricate division of labour, at all levels of human organisation (families, cities, countries, globally). But at all these levels we also compete with other human beings for reproduction and survival, causing diseases in those who are less successful, as the everlasting socioeconomic inequalities in health show.

Exposure to a hostile environment sometimes results from actively seeking novel environments with dangers to which we have not yet developed adequate adjustment mechanisms, or from human interventions that have made the environment more hostile. Novelty seeking may even be an element in our design that has stimulated the spread of the human species around the world—sometimes at the expense of health risks to individual organisms. Human activity has destroyed the local environment in the past, and currently threatens to erode the life supporting aspects of the global environment.

Why do human organisms have design failures? This question can probably best be answered with reference to evolutionary biology. The structural and functional characteristics of the human organism have been shaped over millions of years, and have been carefully selected for their fitness (ability for survival and reproduction) in a hostile environment. Despite its incredible sophistication, however, our design is far from perfect. This lack of perfection is because we have not been created from scratch, but evolved gradually through survival of the fittest in different environments, and that achieving perfection during this process was useless. The only thing that was necessary for us to be alive, was that the genes of our ancestors were transmitted to their children, before their bodies fell apart due to the effects of a hostile environment. Many aspects of aging can neatly be explained from this disposable soma theory, which predicts, among other things, that there must be a limit to further increases in human life expectancy.

Pioneers of Darwinian medicine have suggested that this imperfect evolutionary legacy has a number of specific explanations. One is trade offs or compromises: many so-called design failures contributing to disease have benefits too, for example in other environments or in the same environment at younger ages. Sickle cell anaemia is the classic example: persons heterozygous for the sickle cell gene get substantial protection from malaria. Familial hypercholesterolaemia and other genetically determined cardiovascular risk factors may have provided protection against famine in the not too distant past. Similar trade offs are invoked by the theory of antagonistic pleiotropy, which states that senescence (and disease associated with senescence) is the inevitable product of adaptations that increase fitness earlier in life. For example, mechanisms that limit cellular proliferation protect against cancer in early life, but promote degradation of organ function in later life.

A second explanation of imperfect evolutionary legacies is constraints: some design problems are caused by the fact that evolution is an incremental process. Major changes are difficult, and some elements of our design still reflect their evolutionary biology. The structural and functional design failures are caused by the fact that evolution is an incremental process. Major changes are difficult, and some elements of our design still reflect their evolutionary biology.

Policy implications

While this sketch of a theory suggests that there will ultimately be some hard limits to primary prevention, it also helps in identifying some new approaches to prevention, including interfering with disease mechanisms before detectable disease has occurred, and remediating human organisms’ structural and functional design failures.

Box 2 Possible implications for primary prevention

- Human organisms actively produce disease: by generating signs and symptoms as a reaction to pathological onsets; by engaging in unfavourable exchanges with the external environment; and by creating a hostile abiotic, biotic, and human environment. Whether or not we succeed in primary prevention is, to a large extent, in our own hands.

- There is a enormous theoretical potential for primary prevention of human disease, not all of which has been clearly recognised: (1) interfering with disease mechanisms before detectable disease has occurred; (2) reducing unfavourable exchanges with the external environment, either directly (by limiting these exchanges) or indirectly (by improving the environment); (3) remediating the organism’s structural and functional design failures, either directly (by modifying structure or function) or indirectly (by genetic modification).

- Conventional approaches to primary prevention, which focus on exchanges with the external environment, have not yet been used to the full. Existing knowledge on entry points for reducing unfavourable exchanges with the external environment can be used better. New knowledge will probably emerge from further research into disease mechanisms, and from further research into associations between disease occurrence and environmental and design factors.

- It is probable that we will in the future reach a limit to conventional approaches to primary prevention, because of our dependence on a fundamentally hostile environment. Progress in primary prevention will then become more dependent upon our capacity to interfere with disease mechanisms or to remedy the organism’s many structural and functional design failures.
DISCUSSION AND CONCLUSIONS

My short story of where diseases come from thus reads as follows: In all its manifestations, human disease is a reaction of organisms to, and/or a failure to cope with, one or more unbalancing changes in their internal environments. These are caused by one or more unfavourable exchanges with their external environments and/or failures in the structural and functional design of organisms. In the final analysis, human disease is attributable to the dependence of organisms on a fundamentally hostile external environment and to unfortunate evolutionary legacies.

Although such a concise and broadly applicable formulation can, to the best of my knowledge, not be found in any of the existing textbooks of epidemiology, public health, pathology, or general medicine, most of its elements are of course far from new. It can, for example, be usefully compared with the host agent environment model that has been popular in textbooks of epidemiology and public health. After the near elimination of infectious diseases, whose occurrence had been understood on the basis of a simple host-agent-environment model, many textbooks adopted a similar conceptual model for the understanding of chronic diseases, such as cardiovascular disease and cancer. In addition to infectious agents, the model now included new agents such as nutritive excesses, allergens, and ionising radiation. The latter can only be seen as (active) agents in a metaphorical sense, however, and such a model does not help to clarify the role of behaviours that was relegated to the category of host factors.

More complex models have been proposed, such as the web of causation and more dimensional models taking into account different levels of organisation and different time scales (an eco-social model visualised as a fractal structure, an eco-epidemiological model visualised as a set of Chinese boxes, and a social-ecological perspective visualised as a cube moving through time). The latter are useful frameworks that help to bring more (social) theory into epidemiology and public health. Although these models do not specify where diseases come from, they can be used to structure various strands of explanation, including the type of short story given above, into a larger coherent novel that also tries to explain in some depth how noxae come together in time and space with susceptible organisms.

It may also be usefully compared with the teachings of two giants of public health in the 20th century who have written extensively on this topic, René Dubos and Thomas McKeown. The great microbiologist René Dubos published his influential Mirage of Health in 1959, followed by Man adapting in 1965. In these books he argued that “states of health or disease are the expression of the success or failure experienced by the organism in its efforts to respond adaptively to environmental challenges”. But “complete freedom from disease and from struggle is almost incompatible with the process of living”, because the accomplishments of science and technology “ignore the dynamic process of adaptation to a constantly changing environment that every living organism must face”. In the case of human organisms, some adaptive challenges are due to the fact that mankind follows “urges and strivings that have nothing to do with species survival”, but are entirely the result of novelty seeking. These elements can also be found in the short story developed in this paper, but there are important differences too, particularly because of Dubos’ lack of attention to the role of design failures in disease causation, perhaps because of the less prominent role of aging in population health at the time he worked.

McKeown’s The origins of human disease was published in 1988, and was a bold attempt at classifying all diseases in only three categories: prenatal diseases (all diseases manifested before birth), diseases of poverty (diseases manifested after birth that are attributable to deficiencies or hazards related to lack of the essentials for life), and diseases of affluence (diseases manifested after birth that are attributable to maladaptation or hazards related to industrialisation). This distinction is firmly rooted in an understanding of the links between disease and the material world we live in, particularly the food sources we are dependent on. It seems to over-emphasise the role of too little or too much prosperity in disease causation, however, perhaps as a result of McKeown’s historical analyses of mortality decline in England and Wales that had shown the important role of living standards. While economic circumstances are certainly important determinants of unfavourable exchanges with the external environment, environmental hostilities cannot be reduced to economics. Again, there is little attention to the possible role of design imperfections in disease causation.

Concise generalisations usually come at the expense of being too abstract or too a-specific, and my short story is no exception. Upon arrival at these slightly trivial rock bottoms of disease explanation, some disappointment is perhaps inevitable. It is not more than a sketch asking for further elaboration and refinement, but suggests a few new thoughts about opportunities and limits for primary prevention (listed in box 2). There seems to be a much wider range of opportunities for primary prevention than usually thought, and epidemiology and public health would do well to pay more attention to the opportunities offered by insights in disease mechanisms and the role of design failures.

It is unlikely that we can ever prevent all disease, because our capacity to change the external environment is limited, and so is our capacity to improve the design of the human organism. We are far from having reached these limits, however, as shown by the seven countries study and other exciting results of public health research. Although paradise is not a realistic option, further improvements in our environment seem feasible, and we may in the future also have more opportunities for interfering with disease mechanisms or improving our design. These may contribute to further substantial improvements in healthy life expectancy.

ACKNOWLEDGEMENTS

This paper is based on an essay written on the occasion of New Year’s dinner 2005 of the Department of Public Health of Erasmus Medical Centre, Rotterdam, Netherlands. Useful comments on a previous version of this paper were received from Gabriel ten Veld, Ewout Steyerberg and Jan Vandenburgroeck.

Funding: none.

Conflicts of interest: none declared.

REFERENCES

Lancet causation.

Time of our lives. The science of human aging analysis.

Mortality decline in the Netherlands in the period 1850–1992: a turning point


The triple helix. Gene, organism, environment


Evolution in health and disease

The origins of human disease

The selfish gene

Mirage of health. Utopias, progress and biological change

The age distribution of cancer and a multistage theory of


Critical review of oxidative stress and aging.


Choosing a future for epidemiology: II. From black-box to ecosocial perspective.


Choosing a future for epidemiology: II. From black-box to ecosocial perspective.


Choosing a future for epidemiology: II. From black-box to ecosocial perspective.


Choosing a future for epidemiology: II. From black-box to ecosocial perspective.
