

It is too early to tell what the effect of our knowledge of the human genome will have. There can be no doubt that the environment will still be a major determinant of disease incidence and prevalence but it may be possible to identify differences in personal risk which may benefit individuals. Perhaps the study of those who do not get diseases will become as important as the study of the afflicted.

'Sick individuals and sick populations' is still an important paper and Geoffrey Rose's comments upon the failure of case-control and cohort studies to detect necessary agents which are

homogenous within a population remain entirely valid. However, altering population risk will, for the most part, remain an unreal aspiration.

References

- ¹ Baseline serum cholesterol and treatment effect in the Scandinavian Simvastatin Survival Study(4S). Scandinavian Simvastatin Survival Study Group. *Lancet* 1995;**345**:1274–75.

Commentary: Causes of incidence and causes of cases—a Durkheimian perspective on Rose

S Schwartz and R Diez-Roux

Geoffrey Rose's seminal 1985 article 'Sick Individuals and Sick Populations' and his 1992 book 'The Strategy of Preventive Medicine', have made a huge impact on the fields of epidemiology and public health. A casual Social Sciences Citation Index search yielded over 700 citations of this work. The central lesson that has been integrated into the field is that 'a large number of people at a small risk may give rise to more cases of disease than the small number who are at high risk'.^{1(p.37)} This insight, which has profound implications for intervention and prevention strategies, has been incorporated into research contexts through an understanding of the difference between measures of absolute and relative risk. But there is another aspect to Rose's work that has had a more difficult hearing and that runs counter to mainstream epidemiological approaches solidified under the risk factor paradigm. This is Rose's contention that the causes of cases of disease and the causes of disease incidence may be different and require different types of research strategies. In particular he argues that 'to find the determinants of prevalence and incidence rates, we need to study characteristics of populations, not characteristics of individuals'.^{1(p.34)} This issue has become a central theme in the 'epidemiology wars'² with factions sympathetic to Rose's position arguing that epidemiology has lost its public health relevance because of a myopic concentration on individual-level risk factors.³

Rose's contention that the key to understanding incidence and prevalence lies in 'characteristics of populations and not individuals'^(p.34) is, as Charlton⁴ notes, 'a startling claim'.^(p.607) After all, disease ultimately resides in the individual body and is defined at the individual level. Individual bodies get diseased

and become cases. Population incidence itself is merely the averaging of these individual cases across the population. How is it possible, then, that an understanding of the causes of incidence could be different from an understanding of the causes of cases and, more generally, how can the characteristics of a population enlighten us about disease aetiology?

In order to understand Rose's claim it is essential to examine two key underlying concepts in Rose's writings: the concept of 'cause' and the relationship between wholes and parts. In what follows we will discuss these concepts and then, based on this foundation, indicate five situations where the causes of cases and incidence may deserve distinct treatment.

Two central concepts—cause and the relationship between wholes and parts

Rose's notion of cause

The distinction between Rose's view of causation and that of his critics lies not in the types of factors that can be defined as 'causes', but rather in the criteria used to create a hierarchy among the participants in the causal process. For Rose's critics such as Charlton^{5,6} the most important causes, the ones afforded primacy, are those that define the pathophysiology of a disease. These are the causes that come closest to meeting the standard of Koch's postulates in that they are specific to the disease at hand and found universally, or nearly so, among those with the disease of interest. The priority given to these types of causes is due to the greater scientific certainty and universality with which causal attributions can be made. These causes can be more easily examined with clinical data, manipulated in a laboratory context and are more easily identified in within-population comparisons than more distal, population or social causes. These types of causes, therefore, are given higher

priority, based on the greater certainty about the role they play in the disease aetiology of particular individuals.

Rose, in contrast, develops a hierarchy based on different criteria. He gives priority to more distal causes which he feels hold greater potential for prevention strategies. In Rose's view, these distal causes are often defined at the population level rather than at the individual level, i.e. they are 'exposures' that are characteristics of groups or populations and not characteristics of individuals and are therefore invariant within the group. These causes will not be detected by studies that focus on comparing characteristics of individuals within a population, as many traditional epidemiological studies do. In Rose's words: 'In those circumstances all that these traditional methods do is to find markers of individual susceptibility.'^{1(p.34)} Although Rose does not fully elaborate on exactly what he means by susceptibility factors, the term is used to imply not only genetic susceptibility, but also other individual-level characteristics that lead to disease in the face of the social and physical environmental conditions currently present in the population to which the individual belongs. For example, having an addictive personality would only lead to drug abuse in a social context where drugs were available. This personality feature could be conceptualized as a susceptibility factor that would only be of significance under specific social circumstances. The importance of these susceptibility factors is limited since if the underlying causes 'can be removed, susceptibility ceases to matter'.^{1(p.38)} It is not so much that Rose dismisses these 'susceptibility factors' as causes (indeed they do indeed participate in the disease process and are a primary cause of cases), but the priority is given to population-level causes that facilitate the expression of these susceptibility factors (i.e. allow susceptibility factors to translate into disease) or influence the prevalence or distribution of 'susceptibility' (or individual-level) factors themselves.

For Rose, causal priority is not based on the certainty of the causal attribution or universality of the effect. Rather for Rose, the hierarchy is based in the efficiency with which the removal of a cause could potentially decrease the incidence of the disease—its usefulness for efficient preventative strategies.⁷ Typically, the causes about which aetiological significance is most certain, are those causes which are closer to the level of organization at which the disease is defined and more temporally proximate to the onset of the disease. To use an extreme example for heuristic purposes, a myocardial infarct is caused by a lack of blood circulation to an area of the heart. This is a cause that can be verified with some certainty and is a necessary and universal cause of this disease. On that basis, it is a high-priority cause of myocardial infarct. However, from the perspective of prevention, it is nearly useless as a cause. This cause is so temporally proximate to the damage that the window of opportunity to intervene on this cause is very limited. The closer in the causal chain a factor is to the onset of the disease the less opportunity there is for prevention. Even when intervention is possible, it must be performed at the level of individuals, an inefficient strategy when many individuals are at risk. Of course if a risk factor only appears to be a cause but turns out not to be one, it will have no preventive effect. Causal certainty is clearly of great importance. However, despite the greater uncertainty that might adhere to causes that are more distal from the disease both in terms of level of organization as well as temporality, Rose gives priority to these distal causes because his hierarchy is

based on potential preventative efficiency rather than degree of scientific certainty.

Relationship between wholes and parts

The notion of a hierarchy among causes, however, does not fully explain the distinction that Rose makes between the causes of cases and causes of incidence. Regardless of the priority given to distal causes, should not the causes of incidence merely be the more distal causes of the causes of cases? If so, the causes of incidence and cases would not be distinct.

To answer this question it is useful to examine Rose's understanding of the relationship between wholes and parts—between groups or populations and the individuals of which these populations are comprised. Rose invokes a Durkheimian perspective when he contends that although populations are comprised of individuals, the population has characteristics that are distinct from the mere summation of the characteristics of the individuals in the population. The characteristics of the population may be influenced by characteristics of the individuals but the characteristics and behaviours of the individuals are also shaped by the characteristics of the population.

This relationship can be clearly seen in Durkheim's explication of social facts. Durkheim defines a social fact as 'every way of acting, fixed or not, capable of exercising on the individual an external constraint; ... every way of acting which is general throughout a given society, while at the same time existing in its own right independent of its individual manifestations'.⁸ Social facts include all of the spoken and unspoken rules of society into which individuals within that society are born and educated. The rules have a history that was prior to the history of the individuals affected by them and that is sustained even though the individuals who comprise the group change. They render some things normal and others abnormal, some in the realm of easy choice and some out of reach. One can accept or rebel against these norms but in either case they provide constraints on individual behaviours. Each person is born into a slew of social constraints over which they have limited control. In a similar way the physical environment into which individuals are born exists external to and provides constraints on the individual. It shapes the individual much as the individual shapes the environment. These social and environmental facts interact in a dynamic way with individual-level factors to influence health. While of course these social facts are manifest in the behaviour of individuals, they are distinct from those behaviours and can be usefully examined and manipulated at a level of organization outside of the level of the individual. For example, it has often been the case that laws are changed prior to the change in behaviour of individuals conforming to them.

Durkheim's social facts provide a framework for understanding Rose's contention that the causes of incidence and the causes of cases are distinct. It is based in the conviction that wholes and parts have different characteristics and that therefore, the causes of incidence (the whole) can be distinct from the within-population causes of cases, the parts of which the wholes are comprised. But what lies at the root of this distinction? If incidence is estimated by aggregating numbers of cases over time in a population, how is it that the 'causes of cases' and the 'causes of incidence' may differ?

When we examine the causes of disease, we usually restrict our interest to causes that act within an assumed context—a

tacit causal field—that we accept as a constant background. In examining the causes of physical diseases we typically accept as the ‘tacit causal field’ human biological characteristics that are universal. For example, in trying to ascertain the cause of stroke we usually do not consider as a causal factor the human brain’s functional need for blood flow. This is accepted as the background in which causes of disease function. Given that all humans require blood to flow through their veins that needs to be oxygenated, what is the cause of this organ’s malfunctioning? We therefore look for causes of disease in terms of variations within an accepted (and usually unarticulated) causal field.

When Rose discusses ‘cases’ he appears to mean people within a population who become sick. When we look for the causes of ‘cases’ we refer to those causes that distinguish people within a population and time period who become cases from those within that population who do not. Ubiquitous population characteristics serve as the ‘tacit causal field’. Technically therefore we do not look for all causes of disease but rather those causes that vary between diseased and non-diseased people within that particular population and time period. When one examines individuals (cases) within a population one accepts as the constant given background the social and environmental groups to which these individuals belong. The social and environmental facts of the group are held constant—they provide the tacit field. Since they are held constant they have no detectable influence on the causes of ‘cases’ (i.e. the causes of differences in disease status between individuals within that population).

Moreover social facts (and population and environmental exposures more generally) interact with individual characteristics leading to varied effects. For example, the effects of social facts may differ depending on the characteristics of the individuals, their bodies, and other aspects of the social and physical context—causes referred to by Rose as ‘susceptibilities’. Analogously these individual ‘susceptibilities’ may only contribute to disease in the presence of certain social facts or population exposures. Thus although it may be the population-level exposures that result in the expression of these individual ‘susceptibilities’ as causes of disease, only the effects of the individual-level factors are apparent, the effects of the population-level exposures themselves are hidden. One can see only the susceptibilities (i.e. the characteristics of people for whom the social facts lead to disease) and the particular idiosyncratic ways in which the social facts entered into and interacted with the individual. Only by *looking across groups* can the influence of the social facts themselves be seen.

When one shifts levels of organization and looks at the difference in the rate of the disease between populations or over time, the causes of cases (that is the causes of variation between individuals within a given population) are likely to pale in comparison to the social facts that now vary and can be seen. The tacit causal field is shifted. Differences in laws, customs, physical environments, the shape and extent of social networks, etc., may provide the best explanations for the difference in incidence between groups. The incidence itself, the amount of disease in the population, is a characteristic of the population. It emerges from the interactions among the characteristics of the social and physical environment and the susceptibilities of the individuals within them.

Thus, a key element of the distinction Rose makes between ‘causes of cases’ and ‘causes of incidence’ has to do with the fact

that the causes of within-population variability (causes of cases in Rose’s terminology) may be very different from the causes of between population variability (causes of incidence) which are often population-level or social factors. Studies that focus on causes of within-population variability may thus miss important disease determinants.⁹

It is particularly important to note that the effects of population-level characteristics cannot be simply reduced to the effects of similarly named constructs at the individual level. For example, the effects of living in an area with a high unemployment rate is likely to influence health in many ways other than increasing the probability that an individual would be unemployed. The exposed and unexposed are different at the two levels of organization. At the individual level, the unemployed, (i.e. those without a job) are the exposed while those with a job are unexposed. At the group-level, however, both individuals with and without a job are exposed to the health consequences of living in an area with a high unemployment rate (e.g. stress of job uncertainty, dilapidated housing, interacting with unemployed people, etc.).

We must also be careful not to reify these population-level factors, social facts or ‘causes of incidence’; they must enter the individual body to cause disease. They need to affect individuals and thus ultimately be manifest as causes of individual cases of disease in the more generic sense but not ‘causes of cases’ in Rose’s sense, i.e. they cannot be detected in within-population comparisons. In addition, although population-level factors ultimately cause disease by affecting individuals, they do not necessarily enter the body in a simple causal chain that can be reduced to some particular individual-level factor. Rather, the pathways through which characteristics of populations enter the body are likely to be numerous and interactive. Social and environmental facts, for example, determine proximity to *infectious agents*, influence immune status, and help shape health behaviours. Social facts interact with the specific biological and social history of the individual to shape the particular health manifestation. They influence individual-level factors without being reducible to them. In this way, the disease occurrence in an individual is not just a manifestation of the individual’s characteristics but an interaction between the characteristics of the individual and the environment, both physical and social, to which he/she is exposed and which he/she helps create. Thus the disease of any individual incorporates causes at a level of organization above (and below) the individual.

These aspects of Rose’s work—his hierarchy of the importance of causes and his understanding of the relationship between wholes and parts—are apparent in several different situations that have particular import for the distinction between the causes of incidence and the causes of cases. We discuss each briefly from the least to the most controversial.

Situations in which causes of incidence and causes of cases warrant separate consideration

Causes of incidence as antecedents of particular identified causes of cases

There are situations where, although potent individual-level risk factors have been identified as the causes of a rate increase

over time or place, the key to prevention lies in the social facts related to this change in individual behaviours. In this case, the identification of social facts that may be antecedents to the particular causes of cases may be effective. For example, it is clear that the increase in lung cancer during the second half of the twentieth century in the US was due to an increase in smoking, an individual-level behaviour. One possible approach to disease prevention is to encourage individual patients to stop smoking. However, a more efficient approach may be to examine the cause of the increase in individual's smoking behaviours—the social antecedents to this individual behaviour. When the activities of many individuals within a group change over time, it is likely that social facts play a role. The question of why the proportion of individuals smoking increased may point to potent group-level characteristics (e.g. advertising, role models, promotion of stress reduction) that influence individual's smoking behaviours. Individuals may not be aware of these social antecedents so that asking people why they smoke may not reveal them. In addition, while the smoking behaviours of individuals can be detected as a cause of cancer in within-population studies the influence of advertising, for example, on smoking rates cannot since this exposure is *virtually ubiquitous*. What can be detected is the effect of smoking, which itself results from the interaction between population-level exposures (e.g. advertising and social norms) and individual susceptibility (e.g. psychological characteristics which predispose individuals to addictions and the allure of advertising) but the effect of the population-level exposure itself cannot. Research at a *different level of organization* may be necessary to uncover their effects. Understanding these social antecedents to particular individual risk factors can provide potent and efficient loci for intervention strategies but these antecedents cannot be identified in studies focusing on within-population comparisons.

Inability to detect individual-level factors due to relative ubiquity within a population

There are also situations where individual-level behaviours may provide potent causal explanations for the disease incidence (e.g. why the mean level of a disease relatively high in a particular community) but these individual-level factors may not be identifiable within a population due to insufficient variability. Rose uses the example of fat intake within the US as such a cause. It is possible that individual fat consumption is responsible for the high rates of coronary heart disease within the US. However, studies of individuals within the US may not be able to detect this factor because the variation in fat intake within the US all occurs above the threshold for an effect on this disease. Therefore, fat intake as a causal factor can only be detected through a comparison of fat intake in the US and other countries. As Rose notes, when a behaviour is ubiquitous within a group (e.g. if everyone has a high fat diet) that factor does not distinguish one case from another within the tacit causal field of this society. It may, however, be *very important* in explaining between population differences in disease rates. Any time a risk factor is ubiquitous, it is likely that potent social facts are at work. If everyone (or nearly everyone) has a high fat diet it is likely that there are social norms and other structural factors influencing this behaviour. These ubiquitous individual-level factors, and the social facts behind them, cannot be detected in within-population comparisons.

Contextual effects

Group-level factors may interact with the causes of cases (individual-level factors that distinguish diseased from non-diseased individuals within a group) or may be related to disease independently of 'known' causes of cases. For example, social network characteristics (e.g. number and density of social networks) may have a huge impact on the disease rate controlling for those factors that distinguish diseased from non-diseased individuals within the population. A compromised immune system, for example, may be an important cause of why one person develops an infection and another does not within a population, and the size and density of the social networks within the population may interact with this cause of the cases (compromised immune system) to lead to the particular incidence in the population.

From another perspective, of course, this group-level factor could also be viewed as an antecedent to a different individual-level risk factor (exposure to the viral agent). However, this group-level variable is not reducible to the individual-level variable without loss of information. Knowledge of social networks characteristics may allow a better prediction of future incidence in the group than attempting to do so by determining each and every individual's future risk of exposure to the agent.

There are other contextual effects that cannot be readily reduced to antecedents of any particular cause of a case at all. This is because the contextual effect may lead to a myriad of different causal pathways through which the disease may enter the body, rather than just being the antecedent of a particular risk factor. Neighbourhood socioeconomic environment (which may be related to disease after controlling for the socioeconomic characteristics of individuals) is an example of such a contextual effect. For any particular disease, neighbourhood socioeconomic environment may affect the body through increasing exposure to harmful environmental agents, increasing susceptibility to harmful agents, influencing social network characteristics, prenatal vulnerabilities, etc., which in turn interact with a myriad of different individual-level variables. Contextual effects can only be detected in studies that involve both within-population and between-population comparisons (or both individual-level and population-level factors).¹⁰

Influence of the mean level (incidence or prevalence) of the disease on disease risk

One type of contextual effect, deserves particular consideration—the effect of the mean itself (e.g. incidence or prevalence of disease) on individuals' likelihood of acquiring disease (and consequently on future disease incidence). This is one expression of 'dependent happenings', where the occurrence of disease in one individual is influenced by the occurrence of disease in other individuals.¹¹ This can be seen for both risk factors and diseases. Infectious disease provides the most obvious example where the mean number of infected individuals in a group has an important impact on the number of individuals who will come in contact with the disease agent and therefore the number of people who will acquire the disease. If few people in the population are diseased, then it is unlikely that susceptible individuals will become infected. However, as the rate increases, the probability for any individual of becoming infected increases until the saturation point is reached. Thus, although the disease incidence is calculated from the number of diseased people at a

particular moment in time, this average has an important influence on the disease incidence in the future. A classic example of this type of process is herd immunity, where the individual's risk of acquiring infection depends on the prevalence of immunity in the community. Indeed, the mean level of disease at one moment in time becomes incorporated into the disease experience of individuals at a later moment in time. Thus the dynamic interaction between the group and the individual is embodied in the individual-level expression of the disease. That is, incidence at one point in time is derived in part from past incidence.

Similarly in terms of diseases with behavioural components the mean levels of the outcomes can have an effect on future incidence. The incidence of alcoholism in any society at any particular moment in time is influenced not only by the number of individuals with identified individual-level risk factors such as genetic vulnerability or susceptible personality styles. It is also influenced by societal norms regarding alcohol use and the availability of alcohol in the community. The looser the norms and the greater the availability the higher the probability that a susceptible individual will become an alcoholic. Alcohol norms and availability are, in turn, influenced by the past mean level of alcoholism in the community. In looking at incidence at a given point in time the effect of past incidence of alcoholism may not be apparent. The past incidence provides the current context in which the individual-level factors now operate. However, it becomes apparent that this mean has an effect when future incidence is being predicted. Thus, individual cases (and incidence) at one point in time result in part from the contextual effects of incidence or prevalence of the same outcome in the past. The number of individuals participating in an activity similarly influences the norms about risk behaviours which in turn influences the number and type of people who will engage in them. For example, if smoking is normative, many people will engage in this behaviour without thinking much about it—it becomes a rite of passage. However, people who consider themselves rebels might be less likely to smoke. When smoking is considered deviant, however, it is precisely the rebellious who may engage in the activity. Therefore, such population-level factors may help shape the number of people who are likely to engage in health-damaging behaviours as well as determine who in the population is likely to do so. The effects of past incidence (or prevalence) and norms on risk of disease (or risk of acquiring a risk factor) is another example where the investigation of population-level factors is crucial to understanding the causes of disease, and within-population investigation of the causes of cases (focusing on individual-level causes of differences between individuals) may be insufficient.

Definition of health and illness

Characteristics of populations also influence our very definitions of what is health and what is disease. Rose notes that what we consider normal is influenced by what is prevalent. 'What is common is all right, we presume.'¹(p.32) One implication of this is that social facts may also influence disease incidence in the broadest sense, by determining what we consider to be a disease. Social facts influence our expectations of how many aches and pains are normal, how long we expect to live and what we expect our bodies to look like and our minds to accomplish. Bodily aberrations and biological variants can come to be defined as diseases or redefined as normal. Obesity, intersexed

conditions, senility, acne, post-traumatic stress and gender identity disorder are just a few examples. These expectations change over time and place based on the number of ill people, average life expectancies and other types of rates and norms. These types of influences cannot be understood simply by examining differences between who is and who is not ill within a population.

Rose's contribution cannot be overestimated. The insight that characteristics of populations cannot be reduced to individual characteristics and that both may have important impacts on health suggests new realms for health promotion and disease prevention. But Rose's conceptualization requires a shift in thinking, particularly in contexts where individual autonomy and choice is given great priority. Social facts imply that individual autonomy and choice is constrained by social position and physical environment. One cannot, as an individual, simply choose to be healthy or to behave in a way that increases one's health. There are social limitations to the choices faced by individuals. Rose's conceptualization suggests that although the particular combination of individual-level risk factors of each person with the disease may vary, there are general social and physical factors that may increase the number of individuals who will become diseased. Population-level factors can change the rate of disease in a society even when the prevalence of recognized individual-level risk factors remains unchanged. Which realm is chosen for the prevention of any particular disease is a decision based on the assessed relative efficiency and the potential for unintended consequences of our choices. This, in turn, requires thought and creativity in developing research strategies that can uncover causes at levels of organization other than the individual level. The population level may not always be the best choice for examining aetiology or for intervention. Nonetheless, in many cases it may well be and it deserves recognition and consideration.

References

- Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985;14: 32-38.
- Poole C, Rothman KJ. Our conscientious objection to the epidemiology wars. *J Epidemiol Community Health* 1996;52:613-14.
- McMichael AJ. Prisoners of the proximate: loosening the constraints on epidemiology in an age of change. *Am J Epidemiol* 1999;149:887-97.
- Charlton BG. A critique of Geoffrey Rose's 'population strategy' for preventive medicine. *J Roy Soc Med* 1995;88:607-10.
- Charlton BG. The scope and nature of epidemiology. *J Clin Epidemiol* 1996;49:623-26.
- Charlton BG. Attribution of causation in epidemiology: chain or mosaic? *J Clin Epidemiol* 1996;49:105-07.
- Rose G. *The Strategy of Preventive Medicine*. New York: Oxford University Press, 1992.
- Durkheim E. *The Rules of Sociological Method*. New York: The Free Press, 1964.
- Schwartz S, Carpenter K. The right answer for the wrong question: consequences of type III error for public health research. *Am J Public Health* 1999;89:1175-80.
- Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998;88:216-22.
- Halloran ME. Concepts of infectious disease epidemiology. In Rothman K, Greenland S (eds). *Modern Epidemiology*. Philadelphia: Lippincott-Raven, 1998.