Clemson University
TigerPrints

Publications | Philosophy & Religion

3-2001

A Disease by any other Name: Musings on the Concept of a Genetic Disease

Kelly C. Smith
Clemson University, kcs@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/phil_pubs

Recommended Citation
Please use publisher's recommended citation.

This Article is brought to you for free and open access by the Philosophy & Religion at TigerPrints. It has been accepted for inclusion in Publications by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

This was an invited paper for *The European Journal of Medicine, Philosophy and Health Care, 2001,* 4(1): 19-30.

Scientific Contribution

"A Disease by any other Name:

musings on the concept of a genetic disease"

Kelly C. Smith

Clemson University

Department of Philosophy & Religion

101 Hardin Hall

Clemson, SC 29634

USA

(864) 656-5366 Voice

(864) 656-2858 Fax

kcs@clemson.edu

ABSTRACT

What exactly is a *genetic disease*? For a phrase one hears on a daily basis, there has been surprisingly little analysis of the underlying concept. Medical doctors seem perfectly willing to admit that the etiology of disease is typically complex, with a great many factors interacting to bring about a given condition. On such a view, descriptions of diseases like cancer as *genetic* seem at best highly simplistic, and at worst philosophically indefensible. On the other hand, there is clearly some practical value to be had by classifying diseases according to their *predominant cause* when this can be accomplished in a theoretically satisfactory manner. The question therefore becomes exactly how one should go about selecting a single causal factor among many to explain the presence of disease. When an attempt to defend such *causal selection* is made at all, the standard accounts offered (Koch's postulates, Hill's epidemiological criteria, manipulability) are all clearly inadequate. I propose, however, an epidemiological account of disease causation which walks the fine line between practical applicability and theoretical considerations of causal complexity and attempts to compromise between patient-centered and population-centered concepts of disease. The epidemiological account is the most basic framework consistent with our strongly held intuitions about the causal classification of disease, yet it avoids the difficulties encountered by its competitors.

I. Casual Selection and Causal Pessimism

"For every complex problem, there is a simple, easy to understand, incorrect answer."
- Albert Szent-Gyorgyi

There is a crucial distinction that must be drawn between two related problems in causal analysis. Germond Hesslow (1983, 1984, 1988) does this admirably well in his discussion of causal *connection* vs. causal *selection.* The problem of causal connection is the problem of determining which factors in a complex set, many of which may be correlated to the outcome, are actually *causes* of the outcome. It is this question that takes up the bulk of the scientist's attention, and for obvious reasons. For example, the current debate over whether one's intake of dietary salt increases the risks of coronary disease is a debate about whether and to what extent salt is actually a causal factor (as opposed to merely a correlationally attractive bystander) in the disease etiology.

Casual selection is a fundamentally different problem, however. Causal selection involves identifying which factor(s), among those known to be causally involved in a particular outcome, to cite as *explaining* that outcome. In complex causal situations (which are particularly common in medicine), we simply can not, for practical reasons, cite the entire causal matrix as the appropriate explanation[i]. For example, we would tend to cite the application of a burning match to a pile of hay as *the cause* of a barn fire, without meaning to imply that other factors (presence of oxygen, absence of large quantities of water, etc.) were not also causally involved[ii]. Causal factors other than the one selected (presence of oxygen, etc.) are typically referred to as causal *conditions* (or, derogatorily, as *mere conditions*).

Most human diseases have a highly complex etiology, involving as they do causal factors at all levels: genetic, physiological, systemic, psychological, social, etc. It is thus understandable that, as long as questions of causal connection remain unanswered, considerable effort is expended in an attempt to further elucidate the causal matrix. What we must keep firmly in mind, however, is that describing a disease simply in terms of its etiology is answering a fundamentally different question than that of causal selection. This becomes clear once we realize that, even if all causal connection problems were answered once and for all, causal selection problems would remain. This is because what we regard as an appropriate *explanation* has a pragmatic component that is not captured in causal connection analysis.

Allow me to illustrate this point with a hypothetical example. Suppose an alien physician were to examine a human hospital patient. Suppose further that the alien's technology is so advanced that he is able to determine the precise causal sequence leading to that particular patient's cancer. The alien has thus completely solved the causal connection problem and could, at least in principle, draw a causal map of the development of cancer accurate down to the subatomic level. If you were to ask the alien, "What caused the cancer?" he might simply point to the causal diagram. However, this is unlikely to satisfy us because we are not asking for an exhaustive list of causal connections, but for some *selection* as to which of those factors we should cite as being most *explanatory*.

The alien might be able to satisfy this desire for explanation if some causal factor were either necessary or sufficient for the development of the cancer. For example, if everyone with a particular gene developed this cancer, the gene would be an obvious

choice for causal selection and we could legitimately call this a genetic disease in that sense. We could also select in a relatively straightforward fashion if nobody without some (non-universal) environmental influence developed the cancer. For example, if nobody developed the cancer unless they drank large amounts of cranberry juice – in which case, we might label it an environmental disease. However, there are two basic problems here. First, note that the examples above require information about *populations* of humans. It would not be possible for the alien physician to judge the necessity or sufficiency of the causal factors present in this isolated case without knowing a great deal more about humans than he could reasonably derive from a single patient. Second, given what we know about diseases like cancer, it seems highly unlikely that there will be such a neat identification of sufficient and/or necessary conditions, since the causal matrices are just too complex.

The two philosophers to think most clearly about this problem were J.S. Mill and J.L. Mackie, both of whom ultimately expressed pessimism about the possibility of non-arbitrary causal selection. Mill (1859) points out that, rather than simple necessary or sufficient conditions, what one typically finds are *sets* of conditions, the components of which are sufficient only when considered jointly, with the set itself being non-necessary[iii]. For example, the fire in the barn is the result of a great many factors such as the application of burning match to hay, absence of water, presence of oxygen, etc. – none of which would bring about a fire by themselves (i.e., they are individually insufficient). On the other hand, there are an infinite number of such sets which could bring about a fire in the barn – we might apply white phosphorous to the wall or suddenly

increase the air pressure 100,000,000 fold, for example. Therefore, the particular set of factors in any case, while (jointly) sufficient, is not necessary.

Mill sympathizes with the desire to causally select one factor over another in such situations, but feels such exercises are ultimately arbitrary:

> "The real cause is the whole of these antecedents; and we have, philosophically speaking, no right to give the name of cause to one of them exclusively of the others…Nothing can better show the absence of any scientific ground for the distinction between the cause of a phenomenon and its conditions, than the capricious manner in which we choose to denominate the cause." (Mill 1859, pp 214-215)

Given that most diseases are causally complex, this problem is inescapable when classifying diseases in terms of their etiology, as modern medicine so often does[iv]. The fact of the matter is that much of importance hinges on the causal classification of a disease: social acceptability, funding for research, etc. If such classifications really are fundamentally arbitrary, then we have much to correct.

II. Simplistic Methods of Causal Selection

> "He who would do away with philosophy is the slave of the worst philosophy."
> - Fredrich Engels

Ultimately, *some* of what we mean to capture by classifying diseases as genetic can be salvaged. However, people tend to think about causal selection, when they think about it at all, in highly simplistic terms[v]. In particular, there are three popular approaches which are used in an attempt to settle the causal selection problem in this context, all of which are inadequate.

Perhaps the most common, but also certainly the most hopelessly flawed of these techniques, holds that we are justified in claiming a disease is genetic, provided that genes are *causally involved*. This might be a defensible claim, as we will see, if the analysis went further than this to discuss what kind involvement there is and why it is the most important aspect of the explanation of the disease. However, the typical conclusion of a medical genetics paper does not even begin to consider such things. The recent literature is thus rife with examples of researchers claiming that, because there is good evidence of genetic involvement, the disease is now known to be genetic[vi]. Little more needs to be said to show the absurdity of this than to point out that <u>any</u> disease whatsoever can be classified as genetic on these grounds. All disease must involve, directly or indirectly, genes – if nothing else, genes code for the proteins which make all of biology possible. Even something so obviously non-genetic as lead poisoning, for example, could be called genetic in the sense that some people are likely to have more efficient (gene-based) mechanisms for handling high lead levels and thus will tend not to get the condition when others around them do, etc. We could, with equal justification, argue that all such disease should be classified as *protein based*, since all disease must also involve proteins. Clearly, this kind of analysis does nothing to illuminate the causal situation.

The second approach was first systematically developed by Robert Koch in his set of three key postulates. Since these postulates have done much to shape the current thinking about medical causation, they warrant close examination. Koch argues that we are justified in saying a particular pathogen causes a particular disease whenever three basic conditions are met:

1) The pathogen is always found in individuals with the disease.

2) The pathogen is never found in individuals with conditions other than the disease.

3) The pathogen always produces the disease when introduced into healthy individuals.

Note that postulate #1 is simply a requirement that the pathogen is necessary, while postulates #2 and #3 are requirements that it also be sufficient[vii]. It is certainly true that these postulates served Koch and others well in the early field work for which they became famous. The pathogenic diseases Koch studied are probably uniquely suited to this type of analysis, involving as they do a clearly identifiable infectious agent with well-defined and dramatic onset of a stereotypical set of symptoms. But even for infectious diseases, these postulates are, strictly speaking, too strong. For example, many people are infected with the TB bacillus yet never exhibit the disease, but we do not conclude that TB is not caused by the bacillus[viii]. In any event, infectious diseases are not representative of disease in general and genetic disease in particular.

Consider the case of Cystic Fibrosis, often put forward as a classic example of a genetic disease. As the story is typically told, Cystic Fibrosis is caused by any of several hundred known mutations in the Cystic Fibrosis Transmembrane Receptor (CFTR). Given this account, one might reasonably expect that having a CFTR mutation will lead to CF and, conversely, that lacking a CFTR mutation prevents CF. As often happens in such cases, however, the actual clinical picture that emerges is much more complex:

INSERT TABLE 1

If we apply Koch's postulates to the case of CF, we see the following pattern:

1) Is the genetic anomaly always found in individuals with the disease? **NO**

2) Is the genetic anomaly never found in individuals with other conditions? **NO**

3) Does the genetic anomaly always produce the disease in otherwise healthy individuals? **NO**

According to Koch, therefore, there is no reason to describe CF as a genetic disease – despite the fact that we have a relatively good idea of its etiology and have excellent reason to believe that a gene is a crucial causal player.

There are also epidemiological criteria which are essentially more complex versions of Koch's postulates. Sir Austin Hill's analysis lists eight different criteria which should be considered:

1) Strength: The correlation between the causal factor and the disease should be strong.

2) Consistency: The correlation between causal factor and disease should be observed under varying conditions/individuals.

3) Specificity: The causal factor should be correlated only, or most strongly, with the disease.

4) Temporality: The causal factor should precede the disease in temporal sequence.

5) Biological gradient: The disease should exhibit a dose response curve for the causal factor.

6) Plausibility: There should be a plausible causal story as to how the factor causes the disease.

7) Coherence: The causal story about the factor and disease must cohere with other knowledge we have about other causal factors, etc.

8) Analogy: Similar causal factors should cause similar diseases.

This is undeniably an improvement over Koch's original formulation, since it allows for a much more nuanced description of the causal relationship between a particular factor and the disease. However, this is also something of a drawback: where Koch's postulates will usually yield a clear (if sometimes misleading) answer, Hill's criteria will often not yield much of an answer at all (though they will not as frequently mislead). For example, Hill provides no clear method of *ranking* or *weighting* the various factors. Thus, it is relatively mysterious what we are to make of a putative causal agent which scores well on one criterion and poorly on another. When we examine the suitability of a genetic explanation for CF, we get something like the following pattern:

1) Strength	Unknown[ix]
2) Consistency	No
3) Specificity:	No
4) Temporality:	Yes
5) Biological gradient	Possibly
6) Plausibility	Yes
7) Coherence	Yes
8) Analogy:	Unknown

With three fairly positive indicators, two negative and three uncertain, perhaps one could make a better case on Hill's criteria than on Koch's postulates that CF is actually a genetic disease. Hill's account remains disturbingly vague, however.

III. Manipulability

"A little inaccuracy saves a world of explanation." - C.E. Ayers

There are actually a wide variety of techniques that have been developed to make sense of causal selection[x]. However, the one most pertinent by far to the question of genetic disease is manipulability. The manipulability criterion says that a disease is genetic if and only if it is best controlled (prevented) through manipulation of the genes. This has an obvious appeal for medicine - since the physician's primary interest is in preventing disease, he is likely to advocate (at least implicitly) a manipulability criterion. The basic problem here is that the intuitive appeal of manipulability, like that of Hill's criteria, has been purchased at the price of clarity.

There are two main sorts of ambiguity here. The first concerns how we are to interpret the phrase "best controlled". That is, what exactly makes one form of manipulation *better* than another? Perhaps if one technique is cheaper, less invasive, more permanent, and easier to employ, this will not be a difficult decision. However, how often will the choice actually be this clear? It seems at least plausible, perhaps likely, that within 50 years we will have developed the technology to alter the somatic genes of CF patients and restore them to something like normal health. However, this treatment is likely to be expensive and may carry significant risks. Would the gene therapy be a better manipulation than second generation DNase inhalers, a cheap and effective (if impermanent) treatment? This seems like the kind of decision we would want to let the patient make himself, based on his own personal preferences. Our instincts lead us in this direction precisely because it is very unclear how to decide which treatment is in fact best, or even if there *is* an objective best. If we leave the decision to each patient while maintaining the manipulability criterion, however, CF becomes a genetic disease for patients who prefer the gene therapy and a non-genetic disease for those preferring the

inhaler. This seems a very odd sort of relativity to introduce into our causal classification of a disease.

The second point of ambiguity concerns what kind of manipulability we have in mind. When people speak of manipulability, they are rarely clear as to whether they mean manipulability *in practice* or manipulability *in principle*. Either way, there are difficulties. Suppose we mean manipulability in practice – a disease is genetic if there are or very soon will be effective genetic interventions for individuals with the disease. We would all like to believe that such treatments are just around the corner, but is there really any compelling evidence to this effect? To date, no somatic gene therapy has been successfully vetted in a clinical trial[xi]. Moreover, there are several serious technical difficulties that will have to be resolved before the treatment becomes the standard of care (e.g., targeting all and only the desired cells, etc.). Of course, we might be willing to endorse selective breeding or genetic engineering, where permanent alterations are made to the germ line of the population. This would be more tractable from a purely technical point of view (we have bred agricultural organisms for thousands of years), but these techniques still are not really *practical* because they have been rejected by all virtually all modern countries for compelling ethical reasons. In such a case, the cure literally seems worse than the disease. At the moment, therefore, there are very poor grounds for describing any disease as genetic on the basis of it manipulability in practice. If what we mean by a genetic disease is that it is genetically manipulable in practice, then we will have to refrain from describing any diseases as genetic for a long time to come.

Perhaps what we really mean then is manipulability *in principle* – the question is thus not whether a genetic intervention is feasible now or in the near future, but whether one is

possible on some sort of theoretical grounds. One difficulty that arises immediately is this: what theoretical grounds are we talking about? Since manipulability in practice is something we do not currently have the technology to pull off, the theoretical grounds for a claim of manipulability in practice will have to be very general and vague. I would be willing to admit that, given sufficiently advanced technology, any disease is in principle subject to genetic manipulation. Our hypothetical alien physician, for example, would certainly be able to treat the human cancer patient via genetic intervention, should he choose to do so.

Notice what happens when we make this move, however. First, we are now talking about science fiction, which destroys the original reason physicians favor manipulability: its emphasis on the practical aspects of patient care. Second, with a completely unqualified notion of technological progress, we again run the risk of classifying *all* disease as genetic. One does hear this claim occasionally from medical researchers, but if all disease is genetic, then the description of a disease as genetic is *trivially* true. We could, with equal depth of insight, note that all disease is protein-based.

Finally, even if we are willing to take a very optimistic view of technology and discuss manipulability only in principle, then any given disease could just as easily be described as environmental - the alien physician is also certain to be able to intervene in the patient's cancer by using environmental rather than genetic factors. In fact, it is quite common for putatively genetic diseases to be more amenable to environmental manipulation. Thus, phenylketonuria (PKU) is genetic in principle – it is associated with a defective gene that may one day be repairable. However, there are already dietary interventions that can prevent or mitigate the symptoms and there will likely be more

effective and less intrusive treatments developed along these lines in the future (perhaps a pill containing the missing enzyme so that patients could again ingest phenylalanine, etc.). I grant that this may seem a bit far fetched at the moment, but that is precisely the point.

Manipulability in practice seems to involve either extremely dubious levels of genetic optimism or a willingness to engage in dangerous and far reaching social experimentation. Manipulability in principle seems nothing more than a license to dream about what may come along at some point in the future and is so vague as to produce analyses that are both trivial and non-exclusive. In either case, the original motivation for the manipulability criterion – that it meets the clinical concerns of the physician in a practical fashion – disappears altogether.

IV. The Epidemiological Account

"Metaphysics is nothing more than an extremely obstinate effort to think clearly."
– William James

There must be a way to satisfy both our practical need for disease classification and our theoretical need to do justice to the complexities of causal systems in biology. There must be a causal analysis which captures both the practically-minded focus on individual patients and the necessary elements of population thinking. The proper system would be as clearly decidable as Koch's postulates, but allow for the kind of causal complexity incorporated in Hill's criteria. This system will certainly not be able to classify all human disease, indeed it may fail to clearly classify most human disease, given the complex causal system the human body is. However, such an account would still represent a solid

foundation on which to build. The epidemiological account of disease is an attempt to meet this challenge.

In order to avoid as many difficulties as possible, I set myself a relatively modest goal: I want to provide an account of disease causation which is *minimally adequate* – I make no pretense that this account can not be greatly improved or even that different versions of it may not be appropriate for answering different kinds of questions. I must also make it clear that my goal is primarily to present a normative account of how the phrase "genetic disease" *should* be used, rather than a descriptive account of how it is *actually* used. To be sure, a careful description and categorization of the various uses would be an interesting project, but it is not mine at present. Of course, there is a fine line here - I have no wish to develop a disease concept that is so far from common practice that nobody will ever use it. On the other hand, I do not want to adopt uses of the term which are incompatible or theoretically indefensible.

My epidemiological account is thus designed to satisfy the two most basic and widely held intuitions about disease causation as simply as possible:

1) If a disease is genetic, this must mean that those with the gene are more likely than not to develop the disease[xii]. We might call this the bottom-up or individual causal viewpoint.
2) If a disease is genetic, this must mean that most cases of disease in the population are caused by the gene. We might call this the top-down, or populational causal viewpoint.

These seem entirely unexceptionable intuitions that any adequate account of causal selection will have to incorporate. How could a disease be said to be genetic if those with

the gene are not (in some sense) *likely* to develop the disease? How could a disease be genetic if most people who are its victims *do not have* the gene in question? I thus treat these intuitions as primitive and in need of no further support.

I call my account the epidemiological account because it draws its inspiration from the field of epidemiology. Like epidemiology, it is an analysis of disease that crucially depends on statistical methods applied to populations rather than individuals. It is unlike epidemiology, however, in one crucial particular: while epidemiologists invest a great deal of their effort in finding causes to associate with disease (solving the causal connection problem), my concern is in the explanation of disease (solving the causal selection problem) on the assumption that we have already distinguished causes from conditions[xiii]. To be sure, answering the causal selection question depends crucially on accurate information concerning causal connection. I do not want to appear to underestimate the complexity of this task, but the methods and problems of causal connection analysis have been and will continue to be discussed quite widely in the literature and thus are not the focus of this paper.

Epidemiology is an examination of the properties of populations and thus it is crucial at the outset to specify a well-defined population to which the analysis will apply[xiv]. As we will see, extremely counterintuitive results will be produced if we are unclear or waffle concerning the population to which the analysis applies. Once we have clearly delineated our population and drawn up our list of causal factors, there are a great many questions one might ask. For our purposes, however, there are four which are most crucial[xv]:

1) The Global Question (for the entire population): What is the probability that those who have the gene will contract or have contracted the disease because of that gene?
2) The Diagnostic Question (for someone with both the disease and the gene): What is the probability that the disease was caused by the gene?
3) The Testing Question (for someone with the disease): What is the probability that the disease was caused by the gene?
4) The Prognosis Question (for someone who has the gene): What is the probability that this will cause the disease?

Let's turn to a particular case to develop the analysis we will need to answer each of these questions. Suppose we look carefully at a population of 10,000 people where 12.5% (1250) have a particular disease and 10% (1000) have a gene thought to cause that disease because 80% (800) of those with the gene develop the disease. We then gather the following data about the distribution of the disease and its associated gene:

INSERT TABLE 2

Unfortunately, this table is not accurate enough to answer our questions precisely. This is because some people who do not have the gene nevertheless develop the disease, due to the involvement of other causal factors. Therefore, some of the people who have the gene will develop the disease, not because of the gene, but because of these other factors. Thus, we must distinguish between cases where people have the gene and it causes them to have the disease and those where people have the gene and develop the disease for

other reasons. If we assume that the gene and these other factors act independently, then we can say that approximately the same percentage of gene carriers will develop the disease for other reasons as those in the general population without the gene but with the disease (450/9000 = 5%). We can then move on to the more accurate Table 3:

INSERT TABLE 3

With this correction, we are now in a position to answer, for this population, the original four questions posed:

1) The Global Question (for the entire population): What is the probability that those who have the gene will contract or have contracted the disease because of that gene?

This would simply be the number of individuals whose disease was caused by the gene divided by the size of the entire population (760/10,000 = 7.6%). This would tell us how large a problem the gene is causing in the population and thus, by extension, what could potentially be fixed by genetic manipulation. It does not really tell us anything interesting about whether the disease is genetic, however. We certainly would not want to say, for example, that rare diseases can not be considered genetic (in fact, most paradigm examples of genetic disease are quite rare).

2) The Diagnostic Question (for someone with both the disease and the gene): What is the probability that the disease was caused by the gene?

This may be called the *Simple Etiologic Fraction* (SEF), which in this case would be the number of people whose disease was caused by the gene divided by the total number of people with the gene and the disease (760/800 = 95%). This would tell us how likely it is that a particular patient with the gene will develop a disease because of those genes or, if he already has the disease (and the gene), how likely it is that the condition was caused by the genes. It also indicates what percentage of the population with the gene might potentially be impacted by genetic manipulation to prevent disease.

It is tempting to view a high SEF (above 50%) as an indication that the disease is in fact genetic. However, we are tempted not because of what SEF actually indicates as what we might *think* it indicates. It does not really interest us to know that, in individuals with both the gene and the disease, the gene is or is not usually the causal factor responsible (which is what SEF actually reveals). It would interest us to know that the gene is usually the causal factor amongst those with the disease *in general*, since this would allow us to discover whether the populational causal intuition was met. But this is not what SEF tells us and we can not derive that information from SEF (for this we need to answer the testing question). Simply put, SEF tells us nothing about the percentage of diseased individuals who owe their suffering to their genes. To do that, it would have to incorporate information about individuals with the disease, but who lack the gene (which it does not).

Similarly, SEF tells us nothing at all about the likelihood of developing the disease, given the gene (and thus can not answer the prognosis question). We need this information to decide if the individual causal intuition is met. In order to do that, however, SEF would have to factor in information about people with the gene, but who

remain disease-free (which it does not). Tempting as it might appear on first examination, SEF is pretty useless in answering the causal selection problem.

3) The Testing Question (for someone with the disease): What is the probability the disease was caused by the gene?

This is asking for what epidemiologists call the *Population Etiologic Fraction* (PEF). In this case, it would be the number of individuals whose disease was caused by the genes, divided by the total number of diseased individuals (760/1250 = 61%). This does seem to be getting at something important in our concept of disease intuition. In particular, as long as the PEF > 50%, we know that most cases of disease in the population are in fact caused by the genes. This is precisely the requirement of the populational causal intuition, so it seems we need a stipulation in our concept of genetic disease that the PEF > 50%. We might be tempted to stop here and say that this is the only criterion for genetic status. However, although a high PEF insures that the populational intuition is met, it does not assure us with respect to the individual intuition. Consider the following variation on our original case:

INSERT TABLE 4

Here, although it is quite true that most cases of disease in the population are caused by the genes (PEF > 50%), the gene does not cause most people to develop the disease. This violates our individual causal intuition and thus PEF needs to be supplemented as a concept of genetic disease.

4) The Prognosis Question (for someone who has the gene): What is the probability that this will cause the disease?

Here we are asking for what epidemiologist call *Attributable Risk (*AR). In original case from Table 3, we calculate AR by dividing the number of people whose disease was caused by the gene by the total number of people with the gene (760/1000 = 76%). Again, this does seem to be getting at something important in our concept of disease intuition. As long as the AR > 50%, we know that the gene will cause most of its carriers to develop the disease. This meets the requirement of the individual causal intuition, and thus we must also stipulate in out general account of disease that the AR > 50%. Note, however, that just as a high PEF (answering the populational causal intuition) does not guarantee a high AR, so a high AR (answering the individual causal intuition) does not guarantee a high PEF. Consider the following case:

INSERT TABLE 5

Here, although it is quite true that, in most cases, the disease genes cause the disease (AR > 50%), it is equally true that most cases of disease are not caused by the gene (PEF < 50%). This violates our populational causal intuition and thus we can not use the AR criterion alone.

We can now get down to the business of applying the epidemiological concepts. Given the epidemiological outlook, what exactly constitutes a genetic disease? Clearly, it must involve both PEF and AR, on pain of giving up one of our original intuitions. It remains

an open issue, however, how *strongly* these should be interpreted. One obvious possible answer would be to say that a disease is genetic whenever it is "practically sufficient" within a given population[xvi]:

> **Practically Sufficient (PS):** A disease is genetic whenever the gene's *Attributable Risk* AND *Population Etiologic Fraction* are both 100%[xvii]. In plain English, this means that everyone with the gene has the disease because of that gene AND no one with the disease has the disease because of anything other than the gene.

This accords nicely with our intuitions that being a genetic disease has something important to do with sufficiency and necessity. On the other hand, PS does not make the mistake of claiming that the genes are either sufficient or necessary in the strict sense. Since epidemiological analysis is relativized explicitly to some population of interest, it is entirely possible that what is practically sufficient in one population will not be in another. Practical sufficiency is thus a modification of our ordinary notion of sufficiency to highly complex causal systems - it requires only that that the gene(s) be necessary components of each set of sufficient conditions *which can occur in the population*. PS is thus a more workable empirical notion of causal necessity and sufficiency.

The main difficulty with PS is that, even though it is more practical and restricted than sufficiency and necessity in the strict sense, it is still too strong to apply to the vast majority of human diseases (this would likely be true even if we lowered the threshold value to 95%). Its value for our purposes is thus mainly to anchor the endpoint of epidemiological concepts of disease – it represents as strong a notion of causation as it is

possible to generate using population-relative epidemiological analysis of complex causal systems.

However, the goal is to develop a *minimal* notion of genetic disease. If we are to preserve our two intuitions, any epidemiological concept must include both PEF and AR. But what should the threshold value for each variable be? Clearly, it would have to be at least 50%, as suggested by the use of the term "most" in our original intuitions.

> **Minimally Epidemiological (ME):** A disease is classified as genetic whenever both the Population Etiologic Fraction and the Attributable Risk exceed 50%.

If I refuse to sanction labeling a disease as genetic unless it has at least a 50% PEF and a 50% AR, then I have insured that my intuitions are met. Most people with the disease will have the disease because they have the gene and most people with the gene will develop the disease because of the gene. We could, of course, require that the thresholds be set higher than 50%. While I have no principled objection to this, it will be difficult to defend any number other than 50% as anything other than arbitrary. Moreover, it will be difficult enough to establish that current candidates for genetic disease status are legitimate in a minimal sense.

V. Applying the Epidemiological Concept

"The more you know, the more you know you don't know."

-Aristotle

Let's attempt to fix the endpoints of the epidemiological continuum as it appears in practice. On one end of the continuum we have diseases like obesity[xviii]. Obesity has increasingly been described as a genetic condition, since several genes have recently been

discovered that regulate body weight, at least in mice. In this case, we can safely say that genes are 1) causally involved, 2) in principle manipulable and 3) practically sufficient to induce obesity in certain populations of lab mice. Even in the mouse populations, however, we do not have the data we would need to claim that the condition is genetic in the ME sense. One reason for this, ironically, is that we do not know enough about *healthy* mice. That is, we do not know the prevalence of the "obesity gene" among normal weight mice, and thus can not accurately calculate AR. We also do not know how many diseased mice lack the obesity gene, and thus can not calculate PEF precisely either[xix].

Of course, we could *sample* normal weight mice and diseased mice and at least estimate these values in some rough sense, likely with enough accuracy to pass judgment on the disease's ME genetic status for populations of mice reared under "normal" protocols (e.g., with diet and exercise held constant). However, in human populations, at least in the affluent West, it seems highly unlikely that PEF for genes with respect to obesity will exceed 50%. Although it is certainly interesting to find that genes can induce obesity, these cases are only a small fraction of the causes of obesity. Other factors like diet (which can not be controlled well in human populations, despite the best educational efforts of our medical community) will have extremely high PEF values. Diet will therefore almost certainly be singled out as the explanatory factor, *contra* genes, by any reasonable causal selection scheme. According to the epidemiological account, then, human obesity should not be classed as a genetic disease - genes do seem to play a role here, even an important role, but they are not the predominant cause of obesity.

At the other extreme, a condition like Klinefelter's syndrome seems defensibly genetic[xx]. Klinefelter's is caused by the presence of more than one copy of the X chromosome alongside a Y chromosome and results in numerous problems with the development of sexual characteristics. Here, the genes seem Practically Sufficient (PS) for the trait (AR and PEF of 100%). Again, of course, we will have difficulty supporting this claim as strongly as we would like, since we do not really know if the relative numbers of people with the genetic anomaly who fail to exhibit the condition or who exhibit the condition for reasons other than their genes. However, in the case of entire additional chromosomes and complex symptomologies, we have very sound theoretical grounds to expect that very few people indeed will fall in these categories. At the very least, it seems an excellent bet that Klinefelter's syndrome will qualify as genetic in the ME sense, and most likely in the PS sense as well.

Cystic fibrosis (CF) is a much more complex case. CF is associated with any of at least 300 different known mutations in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene. There is no clear relationship between the severity of symptoms and specific mutations, and there are even cases where individuals with a CFTR mutation do not have the disease(see Table 1). CF thus could not be classified as genetic in the PS sense. Even on the ME account, the case is not perfectly clear. The PEF for the genes (collectively, though not for any single mutation) is probably at or close to 100% - we have, at least arguably, decent data here since it is not unusual for CF sufferers to be tested for the gene. Again, however, the AR is simply not known, though we do know it is definitely < 100%, since we know of individuals with the mutations who exhibit no symptoms of the disease. This certainly should counsel caution in describing CF as

genetic even in the ME sense. However, I would argue that it still seems a good bet (though not an excellent one), given the fundamental function of the CFTR receptor in the cell, that individuals with the mutations who manage to function normally are relatively rare (and thus that AR is fairly high). But they need not be terribly rare - unless these lucky individuals actually *outnumber* those with the gene and the disease, CF will still qualify as genetic in the ME sense (AR > 50%).

VI. Relativity Considerations

> "I am sorry that I have had to leave so many problems unsolved. I always have to make this apology, but the world really is rather puzzling and I cannot help it."
>
> - Bertrand Russell

The epidemiological account is able to handle the complex and highly variable causal systems we know to be operating in human disease, while also restricting itself to questions of a decidable empirical nature. However, this advantageous arrangement does come at a price – as an inherently statistical account, any epidemiological explanation of a disease must be explicitly relativized to some particular population. There is no guarantee that the explanation for a trait in one population will hold true for other populations – indeed, there is often excellent reason to think that it will not[xxi].

The population relativity of epidemiological explanations leads to a number of potential worries. For one thing, it is not clear precisely how one is to choose the population in question. There can be no answer to this other than a pragmatic one - the population chosen must be one of sufficient interest to one's audience. It is true that sometimes this will result in a great many competing explanations for the same trait. While this is certainly imperfect, it is clearly preferable to the alternative of a misleading uniformity.

In any event, it is also true that we very often will have a clearly defined population in mind when we create an explanation. The United States Department of Health, for example, is primarily concerned with the explanation of diseases *within the U.S. population*. Those still worried about relativism creeping into our explanations can take solace from the fact that the population in question must at least be cited explicitly and the rules for what counts as adequate within any population are well-defined. The epidemiological account, therefore, hardly constitutes an "anything goes" abandonment of objectivity.

It also might plausibly be objected that such an account of explanation is completely at a loss when faced with the task of explaining individual occurrences of a trait[xxii]. For example, if a patient insists on knowing what caused *his particular* case of cancer, it does not help him very much to cite the relative prevalence of causal factors within the larger population of which he is a part. This is indeed counter-intuitive, especially given the focus of modern medicine on the care of individual patients. However, before taking this criticism too far, we should consider whether or not there really is an alternative.

Recall the case of our causally omniscient alien physician: he knows everything there is to know about the particular patient's case, but nothing about the more general population. If he really wanted to answer the patient's question about his particular case (other than by simply indicating the entire causal matrix), he would have only two possible routes. First, he could engage in so-called *counterfactual analysis*, where one tries to make projections about what *would have* happened in a particular case *were* the circumstances different from what they actually are. This can become tortuous indeed – as when modal logicians speak of an infinite series of *possible worlds*, similar in some

but not all respects to our own, in an attempt to make sense of the semantics here. For example, we might well wonder what life would be like in a possible world where Hitler postponed the invasion of Soviet Russia until after the fall of Britain. When it comes to causal analysis of disease in individual cases without any information from populations, it is hard to see how this is more than raw speculation of a particularly fanciful kind.

The only other option left to the alien doctor would be to explicitly import populational data. In a population, but not in an individual, there will be variation in causal factors. In a large enough population, one can find almost any combination of relevant casual (risk) factors. By crunching these numbers, we can calculate the actual risk of a particular disease, given any set of initial conditions. In other words, populational data would allow us to solve the problem - *but only by cheating*. We simply can not say much, if anything, about the relative importance of causal factors if we rely only on the perspective of the individual patient – the causal selection problem requires data from populations for its resolution. To suggest, therefore, that reliance on populations is somehow a difficulty is to imply endorsement of an impossible alternative. Without some, at least implicit, appeal to population-level information such as which factors vary and in what way, there are simply no grounds for causal selection, no matter what account one favors.

Thus, an account of causal selection which can explain individual cases is, strictly speaking, impossible. Accounts other than the epidemiological one still use populational information because they must, but they sneak it in the back door without ever being clear about what they are doing. The result is a fuzzy and misleading analysis of the problem. Seen in this light, the epidemiological requirement that the explanation be relativized explicitly to a carefully delimited population is a virtue, not a vice.

VII. Concluding Remarks

The concept of a genetic disease is neither well-developed nor generally defensible as it is employed in the literature. There are very few attempts to make the criteria for causal selection explicit in general, and almost none in the specific case of human disease. One must reason backwards from the kinds of claims one finds in the literature to implicit notions of causal selection, but the notions thus uncovered are not able to withstand the harsh light of critical scrutiny. Rather, they owe their survival to their very ambiguity.

The epidemiological account of genetic traits is an analysis of genetic disease which is both practical and theoretically defensible. This account avoids many of the criticisms leveled against its rivals, while still preserving a use of "genetic trait" which is useful and informative. However, we must apply it with great care. In particular, we must always keep in mind that a great many traits, likely even a large majority, will not meet <u>any</u> defensible criterion of genetic status. This is not a failure of the epidemiological account, so much as an admission that the world of biological causation is far too complex and varied to admit of simplistic categorization.

TABLE 1: PERMUTATIONS OF DIAGNOSTIC CRITERIA AND DISEASE IN CYSTIC FIBROSIS

CLASSIC SYMPTOMS	SWEAT TEST	GENETIC TEST	FREQUENCY
positive	positive	positive	classic CF
positive	positive	negative	occurs
positive	negative	positive	common
positive	negative	negative	common
negative	positive	positive	occurs
negative	positive	negative	occurs
negative	negative	positive	unknown
negative	negative	negative	health

Table 2

	DISEASE	NO DISEASE	TOTAL
DISEASE GENE	800	200	1000
NO DISEASE GENE	450	8550	9000
TOTAL	1250	8750	10000

Table 3

	DISEASE	NO DISEASE	TOTAL
Gene present, causes disease	**760**	**0**	**760**
Gene present, but not cause of disease	**40**	**200**	**240**
NO DISEASE GENE	**450**	**8,550**	**9,000**
TOTAL	**1,250**	**8,750**	**10,000**

Simple Etiologic Fraction = 760/800 = 95%

Population Etiologic Fraction = 760/1,250 = 61%

Attributable Risk 760/1,000 = 76%

Table 4

	DISEASE	NO DISEASE	TOTAL
Gene present, causes disease	**1,000**	**0**	**1,000**
Gene present, but not cause of disease	**100**	**1,050**	**1,150**
NO DISEASE GENE	**700**	**7,150**	**7,850**
TOTAL	**1,800**	**8,200**	**10,000**

Simple Etiologic Fraction = 1,000/1,100 = 91%

Population Etiologic Fraction = 1,000/1,800 = 56%

Attributable Risk = 1,000/2,150 = 47%

Table 5

	DISEASE	NO DISEASE	TOTAL
Gene present, causes disease	**710**	**0**	**710**
Gene present, but not cause of disease	**90**	**200**	**290**
NO DISEASE GENE	**2,000**	**7,000**	**9,000**
TOTAL	**2,800**	**7,200**	**10,000**

Simple Etiologic Fraction = 710/800 = 89%

Population Etiologic Fraction = 710/2,800 = 25%

Attributable Risk = 710/1,000 = 71%

Acknowledgements:

This paper was originally produced for an international symposium on *Complexity and Disease* sponsored by the Dutch Government in conjunction with the University of Copenhagen. I am grateful to all the participants for their insightful comments, and most especially to Henrik Wulff for his warm encouragement and extremely insightful comments that resulted in a paper far better than my original efforts.

References

Brandon, R.N., et. al. (1994) "Sober on Brandon on Screening-Off and the Levels of Selection", Philosophy of Science, 61(3):475-486.

Burian, R. (1981-2) "Human Sociobiology and Genetic Determinism" in The Philosophy Forum, XIII(2-3):43-66.

Collingwood, R.G. (1938) "On the So-called Idea of Causation" in Proceedings of the Aristotelian Society, pp. 85-112.

Ducasse, C.J. (1924) "Causation and the Types of Necessity" in University of Washington Publications in the Social Sciences, 1(2):69-200.

Gardenfors, P. (1980) "A Pragmatic Theory of Explanation" in Philosophy of Science, 47: 404-23.

Gifford, F. (1990) "Genetic Traits" in Biology and Philosophy, 5(3):327-347.

Hart, H.L.A. and A.M. Honore (1959) Causation in the Law, Oxford, Oxford University Press.

Hesslow, G. (1983) "Explaining Differences and Weighting Causes", in Theoria, 49:87-111.

_____ (1984) "What is a Genetic Disease? On the Relative Importance of Causes" in L. Nordenfelt and B.I.B. Lindahl (eds.) Health, Disease, and Causal Explanations in Medicine, Dordrecht, Reidel.

_____ (1988) “The Problem of Causal Selection” in Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality, DJ Hilton (ed.), New York, NYU Press.

Hilton, DJ (1988) “The Problem of Causal Selection” in Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality, DJ Hilton (ed.), New York, NYU Press.

Mackie, J.L. (1965) “Causes and Conditions”, in American Philosophical Quarterly, 2(4): 245-64.

_____ (1974) Cement of the Universe, Oxford, Clarendon Press.

Martin, R. (1978) “Judgments of Contributory Causes and Objectivity”, in Philosophy of Social Science, 8: 173-86.

Mill, J.S. (1859) A System of Logic, London, Longmanns (reprinted 1961).

Nagel, E. (1961) The Structure of Science: Problems in the logic of scientific explanation, New York, HBJ.

Rothman, K.J. (1976) “Causes” in American Journal of Epidemiology, 104:587-592.

Ryle, G. (1949) The Concept of Mind, London, Hutchison.

Salmon, W.(1984) Scientific Explanation and the Causal Structure of the World, Princeton, Princeton University Press.

Smith, K. C. (2000, in press) "What is a Genetic Trait?" in Contemporary Genetic Technology: Scientific, Ethical, and Social Challenges, Krieger Publishing, Melbourne, David Magnus (ed.).

_____ (1998) "Equivocal Notions of Accuracy and Genetic Screening of the General Population" in The Mount Sinai Journal of Medicine, 65(3):178-83.

_____ (1992) “The New Problem of Genetics: A Response to Gifford” in Biology and Philosophy, 7: 331-348.

Sterelny, K., K.C. Smith and M. Dickison (1996) “The Extended Replicator” in Biology and Philosophy, 11:377-403.

Van Fraassen, B.C. (1980) The Scientific Image, Oxford, Oxford University Press.

Wulff, H.R. (1984) "The Causal Basis of the Current Disease Classification" in L. Nordenfelt and B.I.B. Lindahl (eds.) Health, Disease, and Causal Explanations in Medicine, Dordrecht, Reidel.

---

[i] Indeed, one common complaint about the holistic approach to causal explanation is that it becomes very difficult to limit the number of causal factors one is forced to consider, and a kind of "galloping holism" threatens (Sterelney, et. al., 1996). In the extreme, for example, one can make a case that all factors within the light cone of a particular outcome may have to be included in a truly complete causal explanation.

[ii] I revert to well-characterized situations like barn fires when it is important to minimize the intuitive impact of our causal ignorance. One could in principle construct the same sort of example with a disease, but our ignorance of the causal factors and their importance will tend to lead the discussion away from causal selection and towards causal connection issues.

[iii] Mackie (1965, 1974) develops a precise vocabulary to discuss these relationships. The application of fire to hay would be seen as an *INUS factor*: an Insufficient but Unnecessary part of a set of conditions which is Unnecessary but Sufficient for the effect. A set of INUS factors, which is jointly sufficient to bring about an effect, is called a *Minimally Sufficient Condition* (MSC).

[iv] Wulff (1984) offers an interesting discussion of the evolution of disease classification away from symptomology and towards causal agency. It should be clear, however, that if we define a disease in terms of the presence of some particular causal factor, then we have already chosen a particular way to answer the causal selection question. This may work very well for, say, infectious diseases. However, the coming genetic revolution is likely to reveal major flaws in this approach as more and more asymptomatic people with "disease genes" are discovered.

[v] To further complicate matters, questions of genetic disease are tied to misleading and inappropriate conceptions of the nature and significance of genes – though this is not often recognized (see Smith 1999).

[vi] Even in those cases where the researchers are carefully circumspect, reporters covering their work typically are not.

[vii] It may be a bit unfair to show Koch as requiring strict sufficiency and necessity through the use of "always" and "never" in the postulates. Whatever the historical accuracy of this move, it serves to illustrate the poverty of such an approach when applied to complex diseases. At a minimum, Koch's approach would have to be supplemented with another technique (which he does not provide) on pain of leaving the causal selection to the whim of the researcher.

[viii] Indeed, exposure to the TB bacillus used to be essentially universal before modern public health measures were instituted in the Western world. Today, exposure to the bacillus tends to be identified as the cause of TB, but in the 1940's it would be more accurate to identify differences in immunity rather than exposure as the culprit (Stern 1973).

[ix] We have to be very careful not to overestimate the strength of the correlation. It is true that all individuals with Huntington's Chorea, for example, also have the characteristic genetic anomaly. It does not follow from this, however, that all individuals with the genetic anomaly have Huntington's. We simply do not know how many asymptomatic people have the gene in question. What is more, we will not find out with typical testing procedures, which are directed only at those who are sick.

[x] These include abnormality analysis (Hart & Honore 1959, Hilton 1988), unexpected conditions (Gardenfors 1980, van Fraassen 1980), precipitating causes (Ducasse 1924, Ryle 1949), dispositional conditions (Nagel, 1961, Martin 1978), and instrumental efficacy (Collingwood 1938). One of the most promising approaches is to develop a notion of limited or practical sufficiency (Gifford 1990, Wulff 1984b) which I attempt to incorporate in an explicit fashion in my own epidemiological account.

[xi] Worse, there has been at least one unanticipated death which seems to have resulted from gene therapy.

[xii] Actually, this is a bit more complex than it seems here, but I will discuss this in the specific cases later.

[xiii] We must specify an explicit list of causal factors among which we will perform our selection. Should it turn out later that we were unaware of some important causal factor, for example, we will have to redo the analysis.

[xiv] In fact, the reference population must have at least two subpopulations *differing* in the trait in question. This is because the phenomenon actually being explained is the <u>variance</u> in the trait between the two subpopulations, not the trait simplicitur.

[xv] I am indebted to Henrik Wulff for much of the insight behind this categorization and for greatly improving the precision of the numerical analysis which supports them.

[xvi] See also Gifford 1990 and Wulff 1984b for similar attempts to develop a notion of practical sufficiency.

[xvii] Alternately, one might require both numbers to exceed some very high figure less than 100%, say 95%, as we do when calculating statistical significance. This would provide a very strong notion of genetic causation without ruling out every possible exception.

[xviii] Of course, it is an open question as to whether a condition like obesity should really qualify as a disease. It is not my intent to offer a general account of disease, so I will simply assume for the purposes of argument here that it does.

[xix] In fact, this sort of situation poses a major problem for any theoretically adequate notion of disease causation. Studies to determine the actual incidence of disease genes in healthy populations will be extremely expensive and may not be available for many years to come.

[xx] It has been suggested that perhaps Klinefelter's is not a genetic disease at all, since it is not heritable. People's intuitions seem to differ widely on the importance of this point and, in any event, my purpose is to develop a normative account rather than a descriptive one. Unless it can be shown that some particular environmental factor routinely accounts for the genetic changes, the epidemiological account will still describe the condition as genetic.

[xxi] In fact, this will be quite common with some traits: lactose intolerance is a genetic trait for most populations in Western Europe and the U.S. (where consumption of milk products is common), yet it is clearly environmental in many Asian countries. See also Burian (1981-2) and Smith (1992) for a discussion of population-relativity in phenylketonuria, as well as Stern's (1973) analysis of tuberculosis.

[xxii] This is because there is no variation within an individual with respect to the causal factors to be analyzed.