Social and Behavioral Components of Environment in Gene-Environment Studies
John K. Hewitt, Ph.D.

DR. HEWITT: Well, I thank the committee for the invitation to join you and share some thoughts with you.

(Slide.)

My invitation came very recently and so I didn't have a great deal of time to learn about the work of the committee, which I have to confess I wasn't aware of until just a few weeks ago. I read the public comment document since then and have been very impressed with it. But at the time of thinking about this and having been invited, I'm not entirely clear what the kind of audience was, I just thought, well, I better talk a little bit about what I know about and that's behavior genetics, and I'll just say a few things about that.

(Slide.)

Similar to what we just heard, when you look at the—almost any trait, and on this first slide body mass index, IQ is the kind of thing we study, alcohol use, depression, atopy which is going to asthma and eczema and things of that kind, heart rate, cholesterol, personality characteristics. If you look at pairs of identical twins, you are going to find that they correlate very highly or moderately highly for those traits. So we know that individuals who are genetically identical and are being reared in the same household tend to be fairly similar, although they are not actually identical. In fact, there are quite large differences even within that pair. So that suggests that there are environmental influences that differentiate between households in the etiology of all of these different kinds of variables.

If we then go and look at other pairs of twins who are not monozygotic twins and genetically identical but dizygotic twins are no more alike than brothers and sisters genetically, you find that the correlations typically drop down to about half what the monozygotic twin correlations were. Sometimes it's higher and sometimes it's lower but it's often around there. What that tells us if you're raised in the same household but you're not genetically identical, your genes are segregating and you're much less similar than if you're an identical twin. If you put the two together you typically see this kind of relationship.

(Slide.)

When a behavior geneticist looks at those data, he talks about the genetic variation, which I'm not particularly interested in right now, but also the environmental variation. A behavioral geneticist typically starts off by saying, "Well, how much of the environmental variation that's causing differences in the population is shared by members of the family and making them all similar?" And this also includes if there's errors of measurement which are correlated, that would go in there. And how much of this environmental variation isn't shared by members of the families intends to make them different? And that would also include measurement error.

(Slide.)

This slide simply shows, before I pursue that line of thought, the relationship between correlation coefficients that I was just putting up on the board there, on the slide there, and the average difference you'd expect to find for a pair of individuals just chosen from that kind of grouping that had that kind of correlation.

If you just take a random pair of individuals from the population, what you expect for something like IQ, which has a standard deviation of 15, is a typical difference, which in a randomly chosen pair of individuals, one in California and one in New Jersey or wherever it would be, is slightly greater than the standard deviation of the population.

As the correlation rises, the typical difference for that kind of pair of individuals declines but what I really want to show you is something like this: If you have a correlation of about .4 for pairs of individuals, say siblings or individuals in a household, the typical difference—the absolute typical difference of a pair like that is still quite a large quantity compared to randomly chosen individuals in the population.

(Slide.)

So, as I say, behavioral geneticists divide things up into genetic influences and environmental influences, those that are shared by families and those that differentiate members of a family.

(Slide.)

And with some simple assumptions you can come to a conclusion about what's doing what.

(Slide.)

And the slide that was just shown at the beginning of the last presentation looking at the genetic and environmental partition based on twin studies for common disorders does this sort of thing in some sense. Typically, you find a large contribution from the genes for most traits, quantitative traits that we look at like body mass index or IQ or heart rate or personality, and typically you find a contribution of the environment. But what we find our kinds of studies is that a large chunk of that environmental variation differentiates members of the same household. Not all environmental variation is associated with differences between households.

(Slide.)

So individuals who share their family environment, which is indicated by "C" for common, and are genetically identical across the entire genome, that's monozygotic twins, are quite similar but they're not identical by any means.

Individuals who share their family environment but only half of their genes are somewhat similar but often very different.

Environmental differences within families may be as important as environmental differences between families.

(Slide.)

I want to give two brief illustrative examples. The first is body mass index. The second is substance use and abuse.

(Slide.)

Body mass index. This is data from the National Heart, Lung and Blood Institute. A twin study

of obesity. It was published a while back. These are the same pairs of twins followed across their adult life. And these are the correlations for body mass index for identical twins, .8, .73, .72, .69 at the age of 63. These are the correlations for the dizygotic twins, non-identical twins, in that study, .4 something, .4 something, .3 something, .3 something.

There's also information in this study about the continuity of body mass index from one age to the next across the adult life span.

(Slide.)

The result of analyzing those data is that you can estimate the proportion of the variance in body mass index and decompose it into its genetic component and its environmental component, and you can estimate the proportion of that variation which is not shared by members of the twin pair. Now since these are adult twins, it's not surprising that the environmental influences aren't shared but it just makes the point again that if you're looking at environmental influences on a trait of interest, like body mass index, which is the direct underlying quantitative variable for obesity, then you need to take account of the fact that environmental differences between members of the same family are going to be just as important as environmental differences across families.

(Slide.)

And you can take advantage of the information of stability from age to age to decompose the genetic variance into that which is newly arising at any given age and that which is the total, and the same thing with the environmental variance.

(Slide.)

And all this really tells you is—and I'll look at the top first—there are substantial genetic influences on body mass index throughout adulthood and some of the genetic influences in middle age are independent of genetic influences in younger adults.

Individuals leaner for environmental reasons early in adult life are unlikely to sustain their leanness. That's because the environmental variation seems to be arising anew from age to age. So environmental influences may need to be chronic to be influential over long periods for a variable light body mass index.

(Slide.)

The second example is something that we study, which is substance use, abuse and dependence. Substance abuse is clearly a major health issue in this country and so we're interested in that.

(Slide.)

What we find when we look at substance use and the development of dependence, we find that both use and dependence are surprisingly—well, use is clearly common in adolescence but the development of dependence is also quite common in adolescence. By age 18, 70 percent of adolescence are experimenting with alcohol. One in three with marijuana. One in three with tobacco. One in ten with other illicit drugs. And one in five of individuals in our community surveys show dependence on some substance but it's often tobacco. We find few sex differences in substance use or dependence and we find an increase across the adolescent period.

(Slide.)

When we look at family resemblance for identical—MZ twins, who are identical twins, non-identical twins, ordinary siblings and adoptive siblings who are not biologically related, when we look at use of illicit drugs, we see a pattern that's consistent with genetic influences being relatively unimportant but family environmental influences being quite important.

(Slide.)

When we look at something like dependence, which is a much more severe form of the phenotype, which follows use, abuse, interdependence, we find a pattern of variation which is much more consistent with genetic variation.

(Slide.)

So the point is that all aspects of substance use show family similarity. At least some of this is genetic but that's especially so for abuse and dependence.

(Slide.)

And in this slide I simply note the fact that when we're studying behaviors of interest, which will be risk factors for diseases, whether it be substance use, abuse and dependence or dietary factors or exercise factors, we need to be aware that the influence of the environment and the influence of genes may change during the developmental sequence.

(Slide.)

You can't measure the environment one time only and expect to get the answer to whether this is etiologically important or not.

(Slide.)

Well, what kinds of environments do social scientists measure? For this I just want to draw the committee's attention to the national longitudinal study of adolescent health, which is currently funded and lead by NICHD, because these people are social scientists who have enormous experience of measuring social and behavioral environmental influences.

(Slide.)

This slide simply shows the kinds of things that they have focused on. The characteristics of the school, the characteristics of the family, romantic relationships, neighborhood characteristics, community characteristics, peers, and the work environment.

(Slide.)

And on the next slide there are examples of the kinds of things that they've looked at under these general rubrics. The percentage of students who smoke in the school, the students who live with both parents, school cohesion, demographic composition, socioeconomic status. In the family the parents' health, the maltreatment of children as offspring in the family. In the neighborhood, crime, violence, poverty and so forth. And I'll move quickly through those and you have them on your handout.

(Slide.)

The view that I've just expressed is a fairly simply minded view of genes and environment contributing to outcomes. We know that that's way too simplistic a view and we need to worry about things like gene environment correlations, the effects of migration, interactions and so on down this list. Because of the time available, I just want to mention gene-environment correlation especially.

Epidemiological studies that measure environments and correlate them with outcomes are always going to be at risk of drawing invalid inferences if it is the case that the environments are correlated with differential genotypes. One thing that behavior genetics analyses tell us is that the genotype of the individual plays a role in selecting from the available environments so developing a gene environment correlation. This is particularly going to be of interest when you look at diet, nutrition, activity levels and so on.

There is an interaction. You do not impose a diet on an individual. You offer a smorgasbord of opportunities from which the individual chooses. Any study—a large scale population cohort study that's measuring genes and environment needs to make sure that the design and analytic strategy is open to that kind of interpretation.

(Slide.)

I want to finish with some lessons from the Lanarkshire Milk Experiment, which I think is something that every student of experimental or epidemiological design should read about. This was a study conducted in 1930 in the Scottish region of Lanarkshire. It was intended to be a randomized—almost a clinical trial of the efficacy of providing supplemental free milk to children as opposed to not providing it and what was the impact of that on growth.

It was a study of 20,000 children in Scotland and they were—the intention was to randomly assign schools to having a pasteurized or raw milk and then the teachers were to randomly assign the individuals within the school to receive the milk or not, and then the study followed the height and weight growth of those children.

Unfortunately, with the best intentions the randomization was not perfect and there was introduced a confounding between assignment to the milk receiving group and assignment to the group that did not receive milk. Some leeway was left for the individual schools to say, well, we have done the random process, it doesn't look quite right, you haven't got it quite right, we can do some reassignments, and unconsciously there was apparently an assignment of needier children to the milk receiving group, which invalidated the results of the study.

(Slide.)

But what Stuydant (ph) pointed is that his conclusion was that had this study been conducted on 50 pairs of identical twins, one randomly assigned to the milk condition and one not, there would have been equivalent power to detect the nutritious effects of the milk and the experiment could have been much more tightly controlled. He concluded that identical twins are probably better experimental material than is available for feeding experiments carried on other mammals and the error comparison between them may be relied upon to be so small that 50 pairs of these will give more reliable results than the 20,000 with which we've been dealing. That is an overstated case unless the correlation for the outcome measures is very high.

But Stuydant, the statistician, was drawing attention to the very considerable addition advantages of the experimental control that can be exerted than working with a small number of closely matched pairs whose only difference is the experimental treatment.

(Slide.)

In the case of a large cohort study you—we're not going to be assigning MZ pairs randomly to one condition or another but they do still provide the best matched pair control controlling for genotype to look at the effects of the environment. These effects of the environment will be environmental effects that differentiate members of the same family but, as I've said, it appears that these are quite likely to be very important sources of environment.

(Slide.)

So my thought for you to consider as a committee is whether the experimental design of this large—very large population cohort study of genes to environment might not be enhanced by a deliberate systematic sampling of genetically identical pairs of twins who could be assessed probably in much greater detail given the cost efficiency of the small n benefit for a very detailed study of the environment. So using the identical twins to study the environment. That gets around the confounding of the gene-environment correlations that are typically found in large scale region to region differences, for example.

(Slide.)

My conclusions are environmental influences differentiate members of a family as well as making them similar. Gene-environment correlations caused by family genetic variation, migration, environmental selection complicate the interpretation of epidemiological studies of environmental influences.

(Slide.)

The influence of the environment may be different at different stages of development and transitory environments compared to chronic ones may often have transitory effects but, by comparison, genes, of course, are a chronic influence.

Social scientists have long experience of the assessment of environmental influences on behavior and health, and I recommend that you contact someone like Kathie Mullan Harris who is the director of the National Longitudinal Study of Adolescent Health, who is an expert in these kinds of measures.

(Slide.)

And my final recommendation is for you just to consider the possibility of enhancing the design while incorporating a deliberate sampling of MZ twins with extensive assessments of the environment that may provide a powerful efficient well-controlled design for the study of the environmental influences and, incidentally, the study of gene-environment interactions.

Thank you.

DR. TUCKSON: Thank you, both, very much.

And now to lead our discussion, let me turn to our colleague, Julio, who will take us through this.

You're still with us in North Carolina? Are you okay?

DR. SCHWARTZ: Yes, we are.

DR. TUCKSON: You can hear us okay?

DR. SCHWARTZ: I can.

DR. TUCKSON: Thank you.

DR. LICINIO: So the floor is open for questions. Before we—I'd like to start with one question to Dr. Hewitt about how to assess the family in a shared environment because that, I think, is probably the biggest limitation because the toxin-like components, you can always measure them chemically but how do you—how would you suggest in a very large study of a million people assessing shared environment?

DR. HEWITT: Yes. I don't think those are necessarily different kinds of environments. That's a statistical decomposition into, between and within. And how do we assess it? It's assessed statistically in terms of its consequences for similarity. The most direct way, for example, is to look at pairs of individuals who are genetically unrelated that have been reared in the same home to the extent that they correlate positively for a trait of interest that indicates the impact of the environment they've been sharing. How to assess it in terms of what are those environments is presumably then the question that you want, and I don't think those environments are necessarily different. It's not that they are different components of the environment. It's just that they aggregate or differentiate so it's measuring the same environments. The point—my point really is that what we've learned about the extent to which these environments aggregate in the families or differentiate members of the family suggests to us that you will learn a lot about the available environmental range by studying differences between individuals even when they are from the same family.

DR. LICINIO: But let's say in terms of the large study which would be across a million people in several states, how do you measure like psychosocial environmental factors? Are there any kind of batteries of tests you would suggest? What do you assess?

DR. HEWITT: Okay. Well, there are things that are known to be associated with the behaviors that I'm interested in and they are usually characteristics. So parental psychological disorder, parental substance use and abuse, parental style. The treatment of children has turned out to be a very robust predictor of poor behavioral outcomes. So the mistreatment of children. And those things can be assessed by questionnaires and interviews. Both of the parents of children and of the children retrospectively themselves. So that's one class of variables but they are going to be done by interview assessments.

Another class of variables are background variables to do with the school characteristics, the neighborhood characteristics and community characteristics, which can be done at various levels through databases from census tracks and so forth.

DR. SCHWARTZ: I was wondering if I could add to that answer? So one area that we're very interested in exploring is in the Genes and Environment Initiative and the exposure biology program is to develop the biological markers of response to various forms of stress. We intend to

focus on a variety of biological pathways such as oxidant induced stress or specific inflammatory responses or even transcriptional markers as a way of trying to identify how biology is altered by various environments, including behavior.

So while we may not be able to specifically identify the exposure or the behavioral stress, we may be able to identify the biological fingerprint that places an individual at excess risk of developing a disease given a specific genetic susceptibility.

DR. TELFAIR: This is a question for both of the guest speakers. One of the issues, and there is many of course but one of the issues that came up in the work for this project was a utility question and it's a utility question in light of the need to be able to prioritize what the study—what to really focus on given that there may be limitations in resources and time, and that sort of thing.

So I would like to see from your perspective, given that you focused on this, what do you see if you had to make some direct recommendations to focus? What would you think would be critical to look at given your perspective?

Sorry about the feedback.

(Laughter.)

And we're not in Colorado either.

DR. HEWITT: Well, okay, I'll go first since I'm here. It clearly depends on the outcomes which you're most interested in and the interests that I have aren't necessarily the interests that this overall project has. I was just speaking from my experience.

But I have to say that I gave the example of BMI which clearly is something which is related to a wide range of diseases, common diseases, and in that case I would endorse a focus on diet and physical activity, and all of those things that have been characterized as a toxic food environment. That will be an enormously interesting thing, availability of resources for activity and exercise, the saturation of different kinds of fast food versus healthy foods, and so those kinds of things would be enormously interesting and probably productive for the study of obesity and heart problems, diabetes and so on.

It would also be enormously instructive in the context of the kinds of design issues I was talking about to include a focus on individual behavior and individual selection of the environments which are offered to people. I think that would be very helpful so that's one area.

The other kinds of things that I'm interested in are more psychological disorders, substance use and abuse development, and for those things exposure and availability of substances would be very important but they may not be the things that are of most interest to this committee.

DR. TELFAIR: Thank you.

DR. SCHWARTZ: I think in terms of utility there are two components, at least two components that we'd like to affect. One is we have the opportunity by identifying specific environmental exposures that are relevant to the risk of developing disease to decrease the risk of developing disease by altering exposure to those environmental agents. Right now we don't have as much information as we really need to do that.

Second, and I think equally important, I believe that by defining the environmental and genetic risk, we'll be able to understand disease pathogenesis at a much more precise level. By understanding the basic mechanisms of disease, I think we'll have a terrific opportunity to identify disease at a much earlier stage. I think we'll have the opportunity to identify novel mechanisms that may lead to new cures and I think that we'll be able to embark upon interventions of a secondary preventive nature where we know the disease has started but we think that early treatment may avert the full blown development of disease.

DR. TELFAIR: Thank you.

DR. LICINIO: Go ahead, Kevin.

DR. FITZGERALD: Yes. I would like to thank both speakers and ask both of you also in the presentations that you made if you look at other nations that have national databases and things, is there anywhere proof of principle precedent for the kinds of things that you want to do? In other words, have they been done somewhere else to some success, the various sorts of things that you're suggesting we might do?

DR. SCHWARTZ: Well, I can answer that just in the context that there is proof of principle that specific exposures and specific gene changes are associated with the risk of developing disease or the likelihood of preventing disease from occurring. So clearly genetic factors in combination with environmental exposures have been shown to either be protective or detrimental in terms of the development of disease in terms of looking at variations related to TOL receptors that are important in terms of innate immune signaling. It's clear that some individuals with polymorphisms of these receptors are protected from developing asthma that is caused by exposure to these—exposure to bacterial toxins.

It's also clear, as I mentioned in my presentation, that biomarkers of exposures provide a much more precise measure of exposure that allows you to uncover the relationship between exposures and various disease outcomes.

DR. FITZGERALD: I'm sorry. I wasn't very precise in my question. I was more focused on the interaction that you were talking about between your GEI and the large population base studies. Has that been done and what results have come of that?

DR. SCHWARTZ: Not to my knowledge.

DR. HEWITT: But in terms of large scale national studies of pairs of twins, for example, there are a number of good examples in the European countries. Finland has a national registry. Norway, Sweden, Denmark and so on. But I don't think there has been a combination of the large scale population cohort studies with those things on the kind of scale with the kind of opportunity that you're talking about now.

DR. SCHWARTZ: So what I would just add to that is that the tools that are being developed in the GEI and are being tested in the GEI will be directly applicable to a large scale study in the United States as well as large scale studies around the world. And, as I mentioned in the presentation, will be applicable to even small scale studies in terms of other cohorts or other case control studies that are trying to identify the relationship between exposures and genetic variations in terms of the risk of developing disease.

DR. LICINIO: Additional questions?

DR. RANDHAWA: Just one question for my own edification here. So I'm trying to think of the clinical and public utility of the information that we will get and let's take BMI as an example. If you were to get more information that certain individuals with certain genotypes have a 1.5-fold increased risk or a two fold increased risk, what will be doing differently from what we have right now? Would we recommend people to be exercising regularly, eating the right diet? So for people who don't have the high risk genotype, would we say don't do that? And for people who do, are we going to say do it more intensely? What are the other interventions we think we will get once we know this information?

DR. HEWITT: Again, since I am here I'll go. It sounds like a policy issue to me.

(Laughter.)

But presumably at the genetics side, and we were talking about the environment, at the genetics side the intention would be to develop pharmacological interventions that might help control the behaviors that lead to obesity, which are conditional in the genes. That has been the story with leptin as something which has tried to inform the mechanistic understanding that would in turn lead to some changes that one could develop pharmacologically. It hasn't been successful with leptin but that's the intention.

But from the environmental side there is this issue that is unresolved as far as I know as to what it is that you can change in the environment to actually make a difference. At the moment we all know that exercise is—I mean milk is good for children. We know that but trying to demonstrate that it has a specific effect is a different issue.

As far as I know it's not clear. No one has come up with an environmental intervention yet that really makes a difference in the issue of obesity.

DR. SCHWARTZ: That was great, John. I would just add that it's also a matter of—so these healthy lifestyle issues are going to be recommendations regardless of what comes out of the study as long as what we think is healthy remains healthy in terms of these behavioral interventions. But what this study and this approach will do is it will allow us to effectively target populations for more intensive intervention and I think that that is very possible because that is very important because these intensive interventions have been shown to have a big effect on behavior if you look at behavior or if you look at life style changes, if you look at cigarette smoking, if you look at issues related to physical activity. These much more intensive interventions are very effective in altering behavior.

DR. LICINIO: Any additional questions?

I had one question about the size of the study which is supposed to be like one million people. Does it make—especially the environmental factors very difficult to assess? Because like if you go to the point of it's better to do like a smaller study that is well controlled. If you have people from like all over the country it's very different backgrounds and social economic situations. Just in terms of like measuring them in a way that's consistent across these different settings I think is very challenging.

DR. HEWITT: Yes. I mean obviously I was making the argument for enhancing or complementing the large scale study with a more focused study with controlled genotypes in the

case of MZ twins but there's good experience of large national studies, not this large. Again the AD Health Study is one example with a focus on psychosocial, environmental and social environment variables, and it can be done very well.

There's always going to be, as you all know, the compromise between detailed assessments and the number of assessments that you do, which is why it might be worth considering in my view the large study, yes, but also additional components that might provide efficient additional ways of getting at the information.

DR. LICINIO: Even though it's a much more homogeneous—go ahead.

DR. SCHWARTZ: I would also—is it okay if I talk?

DR. LICINIO: Yes.

DR. SCHWARTZ: I would also add that that's why the Genes and Environment Initiative and the exposure biology program within the Genes and Environment Initiative is so important because part of this program is to develop the environmental sensors and the biological response indicators that not only more precisely measure exposure and response to exposure but are tested in larger populations so that we can understand how feasible it is to use them in a large scale study involving a population even as large as a million individuals.

DR. LICINIO: The point I was going to raise before is that there have been these very large birth cohort studies. I'm thinking like right now of the one that was done in the northern most part of Finland and people were followed for many decades. Do you have any kind of—how informative was that in terms of environmental contributions since there was such a controlled set up? Are you familiar with those?

DR. HEWITT: I am not. I don't know.

DR. SCHWARTZ: One of the problems is a lot of these large studies have not collected personal measures of exposure and some don't even have the samples that could be used for personal measures of exposure. I think that study that you're referring to has that problem that it doesn't—it did not assess exposure adequately.

DR. LICINIO: Are there any additional questions or comments?

DR. TUCKSON: I am just running the train here.

DR. LICINIO: Okay. So, if not, I'd like to thank both of you and let Dr. Tuckson continue here.

DR. TUCKSON: Really again on behalf of all of us on the committee, both, David and John, thank you very, very much. You have added significantly to the committee and I'm glad that Julio was able to take us through that.

To let everybody understand the process here: We will take these comments and roll them into the other public solicitation of comments that we are getting, and we will be well informed by all of these perspectives as we go forward.

Again, I remind you all if you have friends, neighbors and relatives that need to comment on the draft report, you've got the web site and all that stuff and you know how to just get them to get those cards and letters coming.

Here's the deal: We convene tomorrow at 8:30. For those who are at the DoubleTree, the shuttle picks us up in the lobby at—what time?

DR. : 7:30.

DR. TUCKSON: 7:30. How come you all know and I don't know?

(Simultaneous discussion.)

DR. TUCKSON: Thank you, Joseph. One for Joe. All right. Pay back, though.

So 7:30 in the lobby.

Now here's the deal. You're on your own for dinner and so we expect you to be on your best behavior because you're representing all of us.

(Laughter.)

See you tomorrow. Thanks again.

(Whereupon, at 5:29 p.m., the proceedings were adjourned.)