

Session Overview and Report from the SACGHS Task Force on Large Population Studies
Huntington F. Willard, Ph.D.

DR. WILLARD: Thank you, Cindy, and good morning. I've been accused of many things. Having a Napoleon complex is not one of them.

(Laughter.)

DR. WILLARD: Nonetheless, here I am.

I want to briefly this morning, before we get into the major session, to review what the Large Population Studies Task Force has been up to and introduce the session today before we hear from our guests.

First to begin and refresh your memory, especially to refresh the memory of those who are on the task force, this is the list of its members. We've convened several times over conference calls with a variety of assigned duties in order to get where we are today, and I thank all the members of the task force, both those who are members of SACGHS and the ex officios.

The issue that was handed to us was to explore the issue of large population studies as one approach to learn more about the relationship among genes, environment and common disease, where the goals of those studies are principally to move towards improvement of health in this country with intermediate steps along the way of determining the mechanisms underlying common and complex disease, and ultimately informing, we hope, treatment and prevention strategies.

What I would like to do in the next 20 minutes is to review the steps this full committee took in assigning the task force, go through the task force's review of a very helpful document that was prepared by an NIH work group, a very comprehensive report that was made available to the public just about the time of our last meeting in June, and then provide an overview for today's session.

So the background is that we were requested back in the very beginning, some two-plus years ago, to weigh in on the value of a large population study in this country. Through our priority-setting process a year and a half ago, we decided that this topic did indeed warrant in-depth study. In October 2004, a year ago now, we formed the task force to guide the committee and to explore the different issues that would need to be tackled, and in the February/March meeting that we had, we spent a full day hearing from presentations that provided us with a number of facts about the nature of large population studies, and specifically looked at some existing projects both within this country and outside of this country in order to give us a sense of what some of the issues would be or could be that would come to the fore.

As part of that, that session ended up facilitating a discussion of a variety of both scientific as well as ethical, legal and social issues around such studies, and we decided at the end of that meeting that the next step would be to develop a report, and today's session is one step along the way towards that report.

The goal would be to identify the key policy issues around a potential large population study mounted in this country and to have the report outline mechanisms that could be used to address the identified issues, and thereby hopefully help the Secretary in his deliberations about exactly how to proceed.

I should say that we were in touch with Dr. Zerhouni's office and this general strategy was endorsed by him and his staff in that we were not going to look specifically, and we won't today look specifically at the scientific issues that underlie a large population study but rather tackle larger and broader policy and process issues that would need to be tackled in any event, regardless of the scientific issues.

So the major action items from our June meeting for the task force were that the task force would review the work group report that came from the NIH and provide an update this morning, which I will do; and then that we would coordinate a meeting to gather input from the scientific and ethics communities, as well as the public at large. However, the task force decided that it really wasn't in our purview or in our particular area of expertise to move towards an in-depth public engagement on this issue, but rather that what we should do is provide the Secretary with our advice about essentially what the best practices were in the area of public engagement and then allow him to decide what the right mechanism and who the right group or groups might be who would engage the public in their support and/or concerns about the nature of a large population study such as this.

So first let me kick off with a review of the work group report. The NIH brought the work group report to our attention at their meeting in March of 2005, although the report wasn't available at that time, and became available just prior to our June meeting. We thought that this report, which is exceedingly comprehensive and was assembled largely through the efforts of NHGRI, with representation from a variety of experts, a very impressive list of experts from around the country, that the best thing that we could do would be to review it as a task force and then present to you, the full committee, our sense of what that task force was all about.

As part of this, I want to bring your attention to a background paper that was prepared by staff that's located in Tab 4 of your briefing books, and this presents the full report of the work group and many of the specific findings that our task force pulled out from that. I'll summarize the highlights here, but our full sense of our review is found in Tab 4.

I believe also in Tab 4 or somewhere in our briefing book is a copy of an article that Francis Collins published in Nature entitled "The Case for a U.S. Prospective Cohort Study of Genes and the Environment," which was very helpful to us, as well as to the scientific community at large, and should be considered part of the record from that perspective.

So I want to review the goals of a potential study as outlined in the work group report, look at the key characteristics of a potential study as outlined in that report, and then examine the key policy issues that were highlighted in that report as well. I will say that there were a number of issues that were raised in the working group report on the policy front. Our task force pulled out the ones that we thought this committee could prioritize, and so I'm not meaning to suggest that everything I mention is everything that was in the original work group report. It was actually a very comprehensive analysis, and we're picking out the ones that we thought were most salient to our efforts here.

So the work group was established to examine the scientific basis for a large population study and examine some of the logistical outlines of such a study, extensive power calculations on what might be gleaned from such a study, the number of individuals who would need to be enrolled in such a study, and what we might expect given the known incidence of different common diseases that are found in our population, and exactly what we might find out from such an analysis. As I mentioned before, this involved a significant cohort of national experts in a variety of fields in genetics, genomics, epidemiology and medicine.

So the goals as put forward in the work group report, the goals of a large population study in this country would be to ascertain and quantify all of the major environmental and genetic causes of common illnesses in this country, and to set a stage for hopefully a future of preventive medicine and personalized health care, and ultimately more effective therapy to address and/or prevent the onset of symptoms in many of these common disorders.

What that study revealed was that probably this would take on the order of half a million to a million participants in a prospective manner in order to look for the development of specific clinical endpoints along the way. This half million to a million participants would need to be sampled from a number of different Census tracts and inevitably would require door to door recruitment over a four-year period.

There was a significant examination of, for example, how and why it would be necessary to over sample individuals from underrepresented minority groups in order to make sure that they were, in fact, well represented in a cohort of this size in order to provide the same level of power for detecting significant trends in minority populations as in majority populations.

The data collection at entry for this half million to a million participants would necessarily include a wide breadth of phenotypes and environmental factors, and of course one can imagine that the largest scientific issue is trying to decide exactly what that list is of the phenotypes that one would wish to collect information on, and the environmental factors one would wish to collect information on in order to then predict outcomes as one goes along.

Necessarily, having started the process, you can't decide three years into the study that I wish we had started collecting information on another phenotype. So the critical step comes at the beginning of such a study.

Yet, all of this has to be balanced versus the expected cost of the project, the potential burden on the individual participants and how much they're willing to tolerate in terms of questions and examinations, and the power calculations of what you actually predict that you'll be able to gain by collecting this information. The conclusion of the work group was that a core group of baseline variables would be collected in all or nearly all of the participants.

Disease outcomes over the course of the study would be assessed using hospital records and outpatient records, as well as other data sources as collected by CMS.

Now, there were a number of key policy issues addressed, and as I said these are the most salient issues that our task force pulled out from the larger working group list, and not the least by any means of which is the nature of public engagement, that a project like this at any magnitude would necessarily require that the public not only be well informed of what the nature of the large population study was to be, but that they also be fully engaged in this and fully supportive of it and feel some sense of engagement and pride in participating in this kind of a project. These same issues have been tackled by other countries that have mounted their own large population studies, some with greater success than others, and that's an issue that in part we want to look at today in terms of the kinds of processes that might be necessary in order to achieve a level of high public engagement.

Another key policy issue concerns the representativeness of the cohort. How do you in a society such as ours, which is exceedingly heterogeneous, how do you begin to determine whether you in fact have a representative cohort? Do you over sample certain groups, as I alluded to a moment

ago, in order to achieve that, and how do you get that balance? Necessarily, this doesn't stand on its own. It feeds back into the public engagement because each of the groups that you would like to sample from will have its own particular issues. The public itself is not homogeneous any more than their genomes are, and so one will have to evaluate public engagement from the perspective of the individual population groups that are being brought into this process.

Clearly, there's a need to examine the issue of collaboration both on an international and national scale. There are projects that are already ongoing, even within HHS, as we heard in our meeting back in March. There are other international large population studies, and the question from a process standpoint that needs to be explored is what kind of collaboration one would have in terms of either data sharing or sharing best practices between these different large population studies.

Access to data in terms of privacy, in terms of who has access, how it will be protected, do the individuals ever get the data back becomes a major issue for any project such as this, regardless of scale, but becomes a particular issue for something that's as dramatically large as this one is, and tied to that is the issue of notifying the results that come from this study back to the individuals and the provision of genetic or genomic counseling in order to ensure that the public is well educated about the nature of the data that might emerge from such a study on an ongoing basis, not in your own little clinic but for a half a million to a million people as one goes through it. That was identified by our task force as a major policy issue.

There are intellectual property issues, as there are with all genome and genetically based research, but particularly a key it seems here in terms of the nature of some of the outcomes and discoveries that might come from such a project; confidentiality and privacy; informed consent, which is a broad issue for all kinds of clinical and translational research, but in this case particularly I think takes on an acute sense of urgency.

The task force also identified the concept of a central IRB to manage a project like this, as opposed to a distributed IRB where each institution that might be involved around the country in trying to collect samples and sign up people into the cohort, that this might be managed much more effectively from a central IRB standpoint, and what are the issues that would be necessary to tackle and solve in getting to that point.

Then we, as everyone does, highlighted the importance of electronic medical records, which is uneven at best in different parts of this country, and yet in this context would need to be brought up to speed in a very significant manner in order to get the best use of the information that one was going to obtain over the course of decades from a half a million to a million enrollees.

So the list of issues that we spent time on came from this review of the working group report from the NIH, as well as the Francis Collins article that I alluded to previously, as well as our own musings and points of view as we discussed in our various conference calls. There's a set, in addition to the policy issues that I've raised here, most of which are on the social and policy end. There's a set of research policy issues as well, some of which we highlighted but most of which were well taken care of by the NIH work group report by itself.

So from this we identified four categories of issues that we thought needed further review and that we as a task force would present this list to the full committee for its consideration today. So let me go through these in kind.

SACGHS Meeting Transcript
October 19-20, 2005

First, broad social issues. Are there data to support the inherent value of such a large population study? Clearly, that's the goal that everyone who considers such a study has, but are there actually data out there that would give us substantial confidence that at the end of the day, or at the end of the decades, that it would be information well worth the effort to emerge from such a study?

Secondly, is a large cohort study the best way to get this information about genetic and environmental influences on common disease, or especially in a heterogeneous population such as we have in this country, and given the nature of some existing cohort studies that are already underway in a variety of different organizations within HHS, are there other ways to approach this issue?

Then lastly, the 900-pound elephant in the corner is the cost of this study. How much will it actually cost, and how does one balance the cost of that study versus other priorities that one has either within HHS in general or within the biomedical research community?

Resource allocation, what tradeoffs would necessarily have to be made if this study were funded. That's always an issue. It's particularly perhaps an acute issue now in the current state of the NIH budget, but nonetheless it's a broad issue that needs to be tackled that we'll need to consider.

Race and genetics, an issue that's been raised before before this committee. Would such a study either increase or decrease the potential stigmatization of individuals belonging to or being assigned to, on the basis of their genome, subgroups of the population, and importantly, would this either reinforce or help dismantle the social constructs of race, and how would one design or consider processes in the design of a large population cohort study that would tip the balance in favor of one or the other of those potential outcomes?

Lastly, from a benefit standpoint, would the benefits of such a study be distributed evenly to all groups within society, or potentially would this exacerbate issues of health disparities rather than address those health disparities?

At the public engagement end, the task force couldn't overemphasize the need for public trust and public engagement and prioritizing the public welfare. How can the public trust, the nature of this kind of project, against a background where their trust of science and genetics in particular, and the government perhaps in particular, is not exactly at an all-time high? How should such a study go about engaging the public both as a single entity and as a number of individual groups, as I alluded to before?

There's also engagement at the level of the scientific community. How can input from the broader scientific community be gathered? The NIH work group obviously engaged a significant number of individuals who contributed to the analysis that was released previously, but there's a much broader scientific community that somehow needs to be heard from in order to either enlist their support or hear their positions on whether such a study is valuable and worth it in the context of resource allocation and the tradeoffs that are necessarily going to need to be addressed across the realm of biomedical research, both basic and translational.

There are a number of access and health care system issues, some of which are issues that this committee has tackled before in terms of their general applicability, but here in the context of a large cohort study, the issues of health disparities, and would the results benefit people who currently have limited access to care, and the issue of diagnosis versus treatment. How will such a project deal with the ethical dilemma that's created by widening a gap potentially between what

can be diagnosed or predicted and what the medical community can actually do something about, a gap that already exists now but would potentially or arguably be substantially widened during the course of such a study?

What is the cost burden to the study participants? These are details necessarily of study design, but on the other hand there are process points here that specifically address the public and would necessarily be part of a public engagement process. How will the cost burden affect access to study participation across the different strata of our population?

How should minority communities be accessed? If the uninsured are part of the study, how will they be accessed? Many, again, are details, but details that the task force felt should be brought to the front in terms of the substantial number of policy issues that need to be tackled.

There are also a series of research issues that the task force limited itself to a consideration from research policy perspectives, not the underlying research basis per se of such a project. How would such a new large population study leverage the existing HHS cohorts that are already underway and are, at least in part, addressing many of the same questions? How can that full leveraging be insured? How will collected samples be secured, stored and disposed of? That's a logistical issue that gets to the issue of public trust and to privacy issues but becomes a process issue that needs to be addressed. And the issue of family member notification, which clearly is relevant to all genetic health issues, and always has been, but in this case, with a half million to a million participants, with substantial amounts of genetic and genomic information being collected about them, to what extent would that information be shared with family members beyond that half million to a million, or not, and what would the processes be that would need to be put in place in order to deal with that particular issue?

A detailed recruitment plan would have to be developed. That is both a policy question and a process question, as well as, of course, a specific issue dealing with the research itself. We felt that guidelines needed to be developed for the application of the research findings and the anticipated technology developments, with particular attention paid to avoiding discrimination and stigmatization as research findings come through over the course of decades.

Necessarily, in a project such as this which would like to look at the interface between the human genome and our environment, we know how to describe the human genome. We're a little less certain about how to actually describe the environment. So from a process standpoint, it's important to determine what the term "environment" means in this context, and how should a variety of environmental, socioeconomic and behavioral variables be measured on one side of the equation to then balance that versus genomic information on the other side of the equation?

We felt it was important to highlight the need for non-coercive recruitment and how protocols for recruitment, enrollment and withdrawal would somehow be kept free of significant incentives that were not somehow coercive or deemed to be coercive by at least some elements of the population. This again is true for all studies but takes on, because of the magnitude of this study, takes on even more importance.

So those were the general policy and social and research policy issues that the task force highlighted, many of which we hope will be informed during today's session and that members of the committee can dig into more deeply, especially in Q&A today. So we thought today's session as we've designed it, the purpose would be to gather input on key policy and process issues and how to address them from various members of both the scientific and the bioethics communities who will be speaking with us today. We also wanted to gather input from experts in the nature of

public engagement to share with us their thoughts on what the best practices are and the variety of mechanisms there are in order to engage the public broadly on this or similar issues.

The purpose of the session is to help us inform the report that we decided we would prepare for the Secretary that would identify for him key policy issues around a potential large population study, and to outline some of the mechanisms that could be used to address those policy issues.

It's equally important, I think, what we're not going to do today, and that is we're not assessing the scientific need for such a study, nor are we assessing the specific scientific aspects of the study or the research design, both because those were anticipated in the work group report from the NIH and because that is not in our immediate purview as the Secretary's Advisory Committee. So although we are not going to be making a recommendation on the need for such a study, or necessarily even the best approach to the study design, the committee may want to identify, as the task force already has, identify this as a policy issue, that there is a specific need to address what the need is and what the study design should be for such a study.

So today's session will consist of three different panels that we'll hear from, representing the three constituencies that I alluded to previously. Each panel will be followed by Q&A, and in the science panel each speaker will be followed by Q&A for discussion and questions coming from the committee at large, and then at the end of the day we will have a full committee discussion in order for us to reflect on what we've heard today and to determine our next steps as the committee decides the nature of the report that we would like to prepare for the Secretary.

So that is the end of my comments. Where are we on time with respect to being over time?

MS. BERRY: Just a little bit over.

DR. WILLARD: So as I said, we're going to hear from initially three distinguished members of the scientific community, and they've been asked to address their perspectives on policy issues surrounding a large population study, and we've asked that they offer us some insights into the best mechanisms and processes for addressing these issues.

We're first going to hear, and we see on the screen, Dr. Gerald Fink via teleconference from MIT in Cambridge. He is a founding member of the Whitehead Institute and is the American Cancer Society Professor of Genetics at MIT. He uses the common baker's yeast to explore critical pathways of cell growth and metabolism. In other words, he's addressing the very issues of gene and environment interactions in yeast that we would need to address in the context of a large population study in human individuals. The applications of his research include cancer research and the development of anti-fungal drugs. He's also intimately familiar with the beginnings of the Human Genome Project and the substantial benefits that the Human Genome Project has brought to the biomedical research community in general, not just those of us who are in the human genetics community.