CCBAR

Chicago Core on Biomarkers in Population-Based Aging Research
The Center on Aging at NORC and the University of Chicago

Chicago Workshop on Biomarkers in Population-Based Health and Aging Research

Sponsored by the
Demography and Economics of Aging Center on Clinical Medicine
June 14 & 15, 2007

National Institute on Aging

center on demography and economics of aging

THE UNIVERSITY OF CHICAGO
MacLean Center for Clinical Medical Ethics

Population Research Center at NORC and the University of Chicago

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The 5th annual Chicago Conference on Biomarker Collection in Population-Based Health and Aging Research was organized and coordinated by Dr. Stacy Tessler Lindau, Dr. Natalia Gavrilova, and Dr. Lis Nielsen, with assistance from Karl Mendoza, Hanna Surawska, Syed Saad Iqbal, Bhairavi Nallanathan, Ada Gomero, and many others.

Financial support for the 2007 Workshop on Biomarkers in Population-Based Health and Aging Research and the Workshop Proceedings came from the Chicago Center on Demography and Economics of Aging through its grant # 2 P30 AG012857, which was awarded by the Office of Behavioral and Social Research, National Institute on Aging. Support also came from the National Institute of Health through its grant # P30 AG012857, the Population Research Center at NORC and the University of Chicago, and the Maclean Center on Clinical and Medical Ethics.

The following Proceedings were prepared from a transcript. Bolded words are defined in the glossary that appears at the end of the proceedings. Please bear in mind the informal nature of the workshop. Participants' presentations and their comments should not be used without permission. Special thanks to Andreea Mihai, Karl Mendoza, Syed Saad Iqbal, Bhairavi Nallanathan, and Ada Gomero for their assistance in the transcript preparation.

© Copyright Center on Demography and Economics of Aging, NORC at the University of Chicago 2008.

ISBN 0-9769978-4-3

Suggested Citation

Lindau S, Gavrilova N (eds). Proceedings of the 2007 Chicago Workshop on Biomarkers in Population-Based Aging Research. NORC - University of Chicago Center on Demography and Economics of Aging. Chicago, IL. 2008.

WORKSHOP SPEAKERS

Andrew Steptoe	PhD
	Research Chair, Department of Epidemiology, University College London
Stacy Tessler Lindau	MD, MAPP
	Assistant Professor, Depts. Ob/Gyn and Medicine-Geriatrics, CCBAR Director, University of Chicago
Arthur Stone	PhD
	Professor, Department of Psychology, Stony Brook University
Tamara Harris	MD, MS
	Chief, Geriatric Epidemiology Section, NIA
Amy Luke	PhD
	Department of Preventive Medicine and Epidemiology, Loyola University
	Medical Center
Thomas McDade	PhD
	Associate Professor, Department of Anthropology, Northwestern University
Lianne Kurina	PhD
	Assistant Professor, Department of Health Studies, University of Chicago
Vilmundur Gudnason	MD, PhD
	Director of the Icelandic Heart Association Research Institute
Joseph Lee	PhD
	Associate Professor of Epidemiology, Mailman School of Public Health
Lainie Ross	MD, PhD
	Carolyn and Matthew Bucksbaum Professor of Clinical Medical Ethics, University of Chicago Associate Director, MacLean Center of Clinical Medical Ethics, University of Chicago
Hugh Hendrie	MB, ChB, DSc
	Professor, Department of Psychiatry, Indiana University School of Medicine
Chandra Reynolds	PhD
	Associate Professor, Department of Psychology, UC Riverside

Noreen Goldman DSc

Professor of Demography and Public Affairs, Princeton University

Arline Geronimus ScD

Professor of Health Behavior and Health Education, University of Michigan

Emma Adam PhD

Associate Professor of Human Development and Social Policy, Northwestern University

Elissa Epel PhD

Assistant Professor, Department of Psychiatry, UCSF

David Weir PhD (University of Michigan, PSC)

Research Professor and Associate Director, Retirement Research Center, University of Michigan

David Laibson PhD (Harvard University)

Professor, Department of Economics, Harvard University

PREFACE

Sponsored by the Chicago Core on Biomarkers in Population-Based Aging Research and the Behavioral and Social Research program at the National Institute on Aging, the 2007 Chicago Workshop on Biomarkers represents an effort to bring together laboratory and population scientists, who study social behavior from a biological perspective. This innovative series of invited workshops grew, initially, from the National Social Life, Health and Aging Project (5R01AG021487), an interdisciplinary study collecting a broad array of biophysiological measures in combination with survey questionnaire data using a home-based, national probability sample, and lay interviewers. Motivated by the enthusiasm and interest of participants attending that first, small workshop, deriving largely from a single project, the annual workshop series has evolved to address the needs and interests of a variety of projects (e.g. Health and Retirement Study; MIDUS; SEBAS; Social Isolation, Loneliness, Health and the Aging Process; AddHealth, NHANES, Canadian Longitudinal Study on Aging, WHO, etc.) and researchers, many of whom receive funding from the National Institute on Aging. Since 2003, the annual Workshop has fostered an expanding interdisciplinary network of senior and junior scientists actively engaged in biomarker collection in population-based health and aging research in North America and Europe.

While growth is one indicator of success, repeat annual attendance from lead researchers demonstrates the ongoing value of the conference. Diverse attendance from disciplines across the social and biomedical sciences, presents a major, unique draw to the workshop. Every year, we see an increasingly broad range of attendees from across the social sciences (e.g. sociology, anthropology, economics, public policy, political science, demography, psychology) and both clinical and basic science biomedicine (e.g. pediatrics, ob/gyn, internal medicine, geriatrics, otolaryngology, dermatology, infectious disease, cardiology, neuroscience, epidemiology). We have also engaged bioethics participants, as a consequence of an ongoing CCBAR collaboration with the University of Chicago MacLean Center on Clinical Medical Ethics. This year, we took an international approach with a keynote talk by UK professor Andrew Steptoe and participants from Iceland and all over the United States. Each speaker provoked thought and lively discussion. These memorable sessions have been positively reviewed and have resulted in translation of technology, materials, and ideas from fields typically beyond the scope of most academic researchers.

Each year, the Chicago Core on Biomarkers in Population-Based Aging Research (CCBAR) at the NIA NORC-University of Chicago Center on Demography and Economics of Aging, publishes the Proceedings of these workshops. They are distributed to Workshop attendees and NIH colleagues and are posted on the CCBAR website at http://biomarkers.uchicago.edu. We have found that these Proceedings provide an excellent reference, particularly for investigators new to this field. We thank each of the individuals, who shared and edited their presentation for publication, and for the very engaged participants, whose repeat attendance and intellectual involvement, fuel our enthusiasm for continuing this Workshop series.

Stacy Tessler Lindau, MD, MAPP
Lisbeth Nielsen, PhD

Introduction and Welcome: **Lis Nielsen, PhD**

Keynote Address: "***Psychophysiological processes underlying emotional triggering of acute coronary syndromes"***

Andrew Steptoe, PhD (University College London)

Thank you very much, and thank you, Linda, for that kind introduction. I'm delighted to be here and to have the opportunity to come to this workshop, which I haven't been to before. But I've heard about it, and it sounds like a very interesting occasion.

Good morning. I'm Andrew Steptoe from University College, London. I'm a Professor of Psychology working in the Department of Epidemiology and Public Health. I worked for a number of years in a Medical School in south London, working on psychological aspects of physical illness. About five or six years ago, I moved to UCL to join Michael Marmot, who is the head of the Department of Epidemiology, to carry out more intensive biological research in nested studies within the Whitehall II cohort.

"And secondly, because he famously said, 'My life is at the mercy of any rascal who chooses to put me in a passion.'"

I'm going to be talking about a fairly specific issue. As you can see, I've been given quite a long title here: "Psychophysiological processes underlying emotional triggering of acute coronary syndromes." Although it's a rather specific issue, it does, I think, illustrate a number of the themes which are relevant to this workshop, including issues such as the choice of biomarkers, the settings in which they are assessed, the integration of different approaches -- clinical, population, and laboratory approaches. So I hope even those people who aren't particularly interested in acute coronary syndromes will get something from this.

We do have a fair length of time for this presentation because of the change in the format of the morning, so please feel free to interrupt if you would like to ask things as I go along.

Let me just outline for you, for those who are unfamiliar with this area and what one means by an acute coronary syndrome. Although it's a fairly common term in clinical practice, it's less used in research, where people still continue to work specifically on myocardial infarction. Myocardial infarction is a major component of acute coronary syndrome, but this term is now used to embrace a broader constellation of syndromes which have a rather similar etiology. It includes not only the traditional MI, typically an ST-elevation myocardial infarction, but also non-ST-elevation MI, and unstable angina.

My interest in this area really began with this man here - John Hunter, who was an 18th century physician, one of the founders of medical anatomy, and a person who was particularly interested in integrating structure and function in biology. I was interested in him for two reasons. First of all, he worked at St. George's Hospital in London, and that's where I was born. Indeed, that window up there is where I spent the first night of my life. It has changed slightly. I think the ambulances have got a bit better than that. But he was one of the founders of the Medical School.

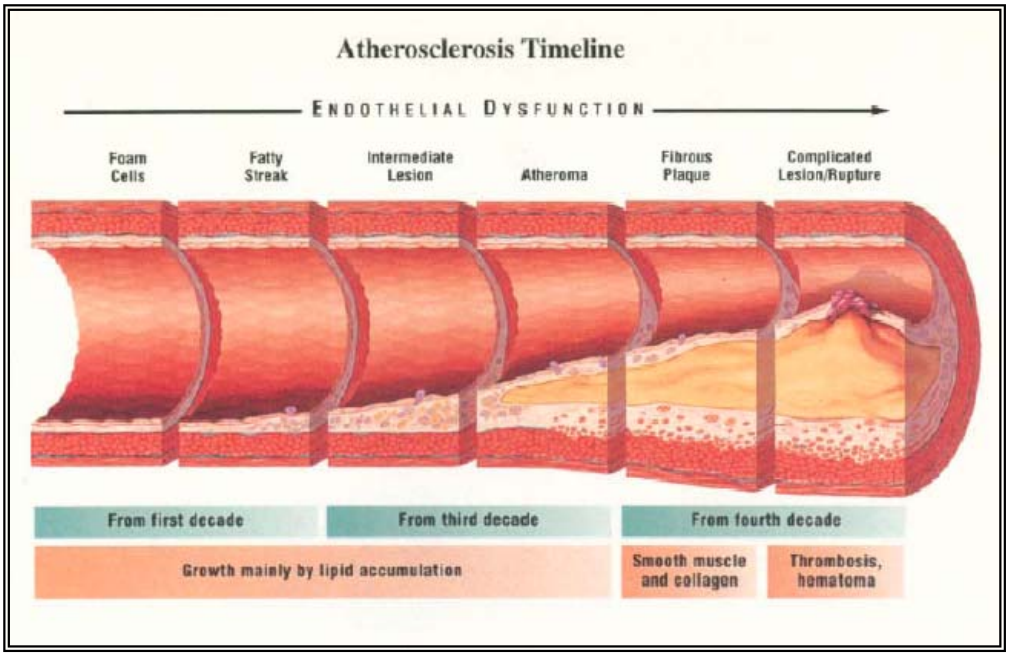
And secondly, because he famously said, "My life is at the mercy of any rascal who chooses to put me in a passion." He was aware of his vulnerability and one morning in 1793, he went to work as usual and went to a board meeting at the Medical School, where he had a row with his colleagues about student fees, and collapsed with chest pain and died in the sofa, which is still to be found in the library at St. George's. I worked at this same Medical School for a number of years, and was Chair of the Academic Board, and I can quite understand how it might drive one to an acute coronary syndrome.

He and his brother were both famous anatomists, so, of course, his brother anatomized him straightaway after he died, and found that he had advanced coronary artery disease, and very, very sclerosed coronary arteries. It's still generally true that people who have acute coronary syndromes and acute cardiac events do have advanced heart disease.

Many of you will be familiar with this cartoon showing the progression of coronary arthrosclerosis from early life through to middle age and older age, which are the times when clinical events take place. However, there's

been quite a change over the last fifteen years in understanding what happens with acute coronary syndromes, and the old idea that the arteries simply got more and more blocked, and finally got clogged up with a thrombus forming, doesn't seem quite accurate anymore.

There are three phenomena which really suggest that something different is going on, and I summarize them in the next slide. First of all, the severest lesions, if you look at them using techniques such as intravascular ultrasound, are not necessarily the ones that are at highest risk of rupture. So it's not just in the parts of the arteries where there's the greatest stenosis or blockage that you're likely to get an acute event. There's something different about the plaque in the areas that are going to rupture. In addition, episodic plaque rupture is a relatively common phenomenon and only occasionally provokes an acute coronary syndrome. So when you look at people's arteries, what you find is that there's a whole host of healed ruptures which can be observed in the scan, so there must be some other factors which contribute to whether or not a cardiac event, or intravascular event, actually leads to a clinical event.



The next slide shows you the cross-section of a fatal acute coronary syndrome. In the centre is the lumen of the blood vessel, which is completely filled with a thrombus. Here is the ruptured plaque. You can see the cartoon of this fatal cardiac event on the same slide. Here is the lipid-filled pool of atheroma. And there has been a rupture at this point here, and the formation of a thrombus, and then the blockage of the artery.

When you look at acute coronary syndromes in this way, there are actually three different phenomena that are required for this type of advanced disease to turn into a clinical problem. One is what we might call a hemodynamic phenomenon; that is, to do with vasoconstriction and with sheer stress forces on the vessel wall that will lead to a potentially vulnerable plaque rupturing. The second is a local inflammatory response, which is associated with changes in the blood – specifically with a pro-thrombotic blood environment, high levels of platelet activation and other clotting factors.

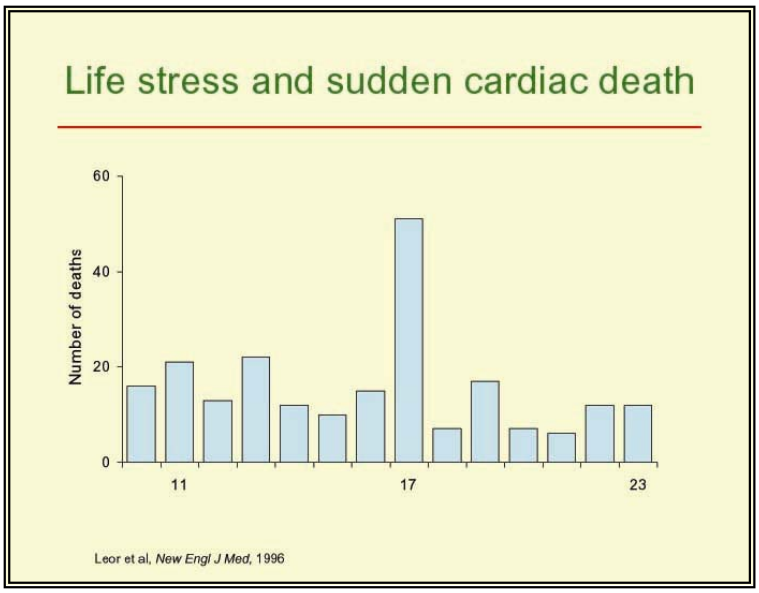
There appear to be a whole series of rather acute and short-term clinical phenomena which have to come together along with the long-term risk. This has led to the notion that there are acute triggers of cardiac events superimposed on the long-term developmental factors, and on the long-term psychosocial influences with which many of us are familiar. This slide shows a typical definition of an acute trigger: a stimulus that can be internal or external, that provokes relatively short-term changes which lead directly to the onset of pathology. Typically, triggers are studied in the one to two hours before the onset of symptoms.

What I'd like to do this morning is to talk about two things: first of all, to mention rather briefly the evidence that emotional factors can trigger acute cardiac events; secondly, to focus on what psychophysiological processes are involved and how we can study them.

Many of you here will be familiar with the general literature relating psychosocial risk factors with the development of coronary heart disease, and, indeed, many participants in this Workshop have contributed to this literature. There's a general consensus, I think, of a series of factors which seem to be associated with the long-term development or acceleration of coronary

heart disease, including low socioeconomic position, various chronic stress factors – in particular, factors in the work environment have been studied most extensively – coupled with an impoverished social environment, and factors such as social isolation; and in addition, certain psychological characteristics, of which depression has been studied most completely over the last decade or so. Anger and hostility may also be important.

These factors are primarily studied in relation to the long-term development of arthrosclerosis. That is, one takes disease-free populations, studies these exposures, and then one follows people up longitudinally to see which people succumb to disease. They're not necessarily the same sorts of factors which are going to be involved in the triggering of the endpoints of acute cardiac events, because many of these may not be particularly acute experiences. And so we may have to look for slightly different types of psychosocial factors in order to understand what goes on at this stage of the disease process.



There are two broad ways to study these acute emotional triggers: The first method is the population-based method, which is to look at various types of public events, and the incidence of acute coronary syndromes following those events. These are events which can be timed very precisely. What we can do is to look before and afterwards at whether there is an increased incidence of acute coronary syndromes in the period following that event.

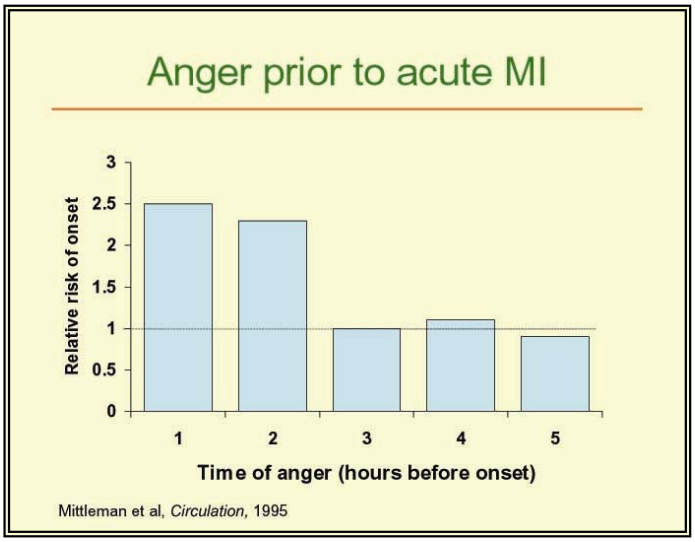
One of the best studies using this approach was carried out on the Northridge Earthquake, which took place on the 17th of January, 1984, in the northern Los Angeles area. This study is a classic in the field because it was possible to really take a population approach to the incidence of sudden cardiac death and acute myocardial infarction. And this slide here shows the proportion of people in that area who died of acute coronary deaths on the day of that earthquake, compared with the days before and afterwards, and various other control periods in the preceding and succeeding years.

There continue to be studies of this nature. In particular, studying sporting events is very popular type of investigation in Europe because fans get very would up during matches. What's been found in at least some studies is that when people watch football matches and then their teams lose, there is an increased rate of death or myocardial infarction. But it's quite difficult to work out whether these are really emotionally triggered events, because as you can imagine, when people watch football, they tend to drink a lot and they're often standing up in hot environments. There are all sorts of factors that could be contributing quite apart from the emotional upset of finding your team not doing very well.

Most cardiac events, of course, don't take place around the time of some major public trauma, such as an earthquake or a terrorist act. So most of the literature in this area is based on interview studies, talking to survivors or to their relatives about personal experiences around the time of symptom onset. The earliest one of these was the study shown in the next slide of around 850 patients who were interviewed following admission to hospital with myocardial infarction. The slide shows the proportion of people who reported different types of possible trigger of infarction.

There are two major problems with this type of work. The first is what I call here "recall errors and retrospective bias" -- that people may not actually remember what happened in the period immediately preceding the onset of symptoms, and they may retrospectively reinterpret their experience based on the fact that they have now had a heart attack. This problem can be got around to some extent by a much more systematic approach to interviewing people and talking to them about exactly what was happening, rather than what they were particularly feeling or what they think caused their cardiac event.

The other problem is the base-rate problem: the sorts of experiences people report are commonplace in their lives, and may not be very specific to the period



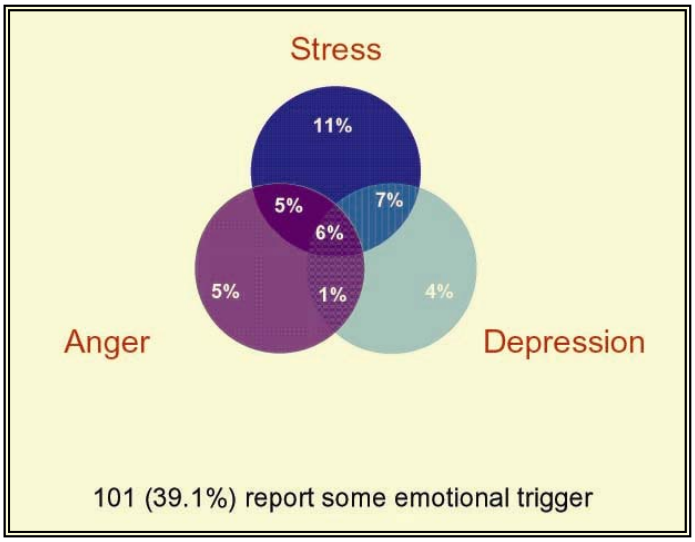
immediately preceding the cardiac event. Something like smoking, for example, is difficult to see as an acute trigger because smokers tend to smoke very frequently, and there may be nothing to distinguish the occasions when smoking is followed
by and acute cardiac event compared with times when it isn't.

It's because of this that the case crossover method of statistical analysis has been developed specifically to look at these time-limited types of phenomenon. This is a kind of case-control method, but using individuals themselves as their own controls, with two or more comparison periods of time as the control periods. So you have a control period, which might be the one or two hours before the onset of symptoms, and then comparison periods, for example -- the same time of day twenty-four hours later, or the typical rates of exposure to the typical type of challenge. Using that sort of methodology, these are the factors which have emerged as acute triggers of cardiac events: physical exertion, sexual activity, marijuana and cocaine, emotional upset, and acute anger.

We've published a couple of systematic reviews of that literature for those who are interested. What this literature suggests is that these factors do seem to be associated with an increased incidence of the onset of symptoms of an acute coronary syndrome. One of the best examples is from the Onset Study, from a group at Harvard, looking at the relative risk of acute myocardial infarction following a period of extreme anger. What this slide shows is the relative risk or likelihood of anger being followed by an acute cardiac event. Anger in the one to two hours prior to onset seems to be associated with increased risk.

I would like show you some details of a study we've carried out in this area just to illustrate to you the sorts of methods that are typically used. This was a study of just under 300 patients who were admitted with acute coronary syndromes, with about an eighty percent response rate. What we did in the study was to interview people an average of two to three days after admission using a structured interview focusing particularly on the two-hour period prior to symptom onset, and comparing this with two control periods in the case crossover method -- the same hours, twenty-four hours earlier, and then an estimate of exposure over the previous six months. So if, for example, an acute myocardial infarction started at three o'clock in the afternoon on a Thursday, then one would be ask people about one to three o'clock in the afternoon on that Thursday -- that's the two hours prior to onset -- and then also one to three o'clock in the afternoon on the Wednesday as well, and then compare their experience over these two time periods.

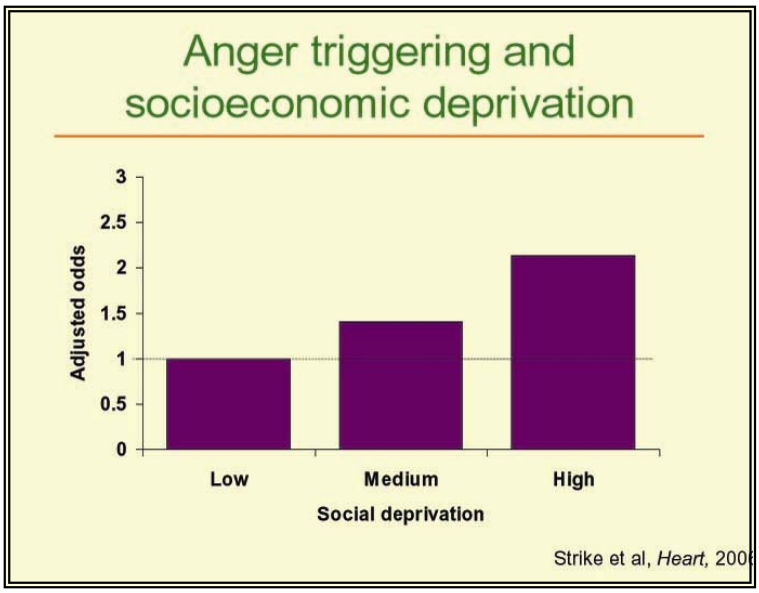
Here is a breakdown of the number of people reporting some kind of emotion during that period, and it's quite high. About forty percent of people



report some kind of severe emotional experience during that two-hour period. But, of course, some of these people also report the same types of experiences during the control periods, as well. But using the case crossover technique, what we found was a very similar result to the one produced in the Onset Study; that is,

an odds ratio of just over two for a period of anger being followed by the onset of acute coronary syndrome, with an incidence of around fifteen percent. Interestingly, this pattern was more common in people of lower socioeconomic status. This slide shows the adjusted odds in relation to social deprivation. This is the high-deprivation group, and this is the low-deprivation group. Anger triggering was also more common in younger rather than older patients. It seems to be a trend in the literature that slightly younger patients – under 65 or so – tend to be more likely to experience these kinds of emotional trigger.

We also found an interesting association with acute sadness, which hasn't been studied before. We found that severe levels of acute depressed mood, stimulated by such things as a recollection of an anniversary, or an interaction with somebody that led to high levels of acute sadness, which was slightly independent of anger, were associated with increased risk. Now these methods are not perfect by any means, because they are all dependent on retrospective report, but they at least give you an idea of how one goes about trying to assess emotional triggering.



What I'd like to do now is to focus on what psychophysiological processes are involved, and how we can study these in the contexts both of a clinical problem and of population patterns of psychobiological response.

Knowledge about the biological factors that potentially link psychosocial experience with coronary heart disease has really expanded a great deal over the last few years (Slide). In epidemiological studies, the early studies tended to focus on traditional measures such as blood pressure or cholesterol levels. But more recently there's been a real expansion in study of other types of biological response – in particular, inflammatory responses. So there's been increased interest in looking at fibrinogen, CRP, IL-6, and heat shock proteins, a particular interest of mine, studying endothelial function, various aspects of the clotting processes, and looking at autonomic dysregulation through such measures as heart rate variability.

There's a growing literature relating these factors with a low socio-economic position, with work stress -- high demands/low control -- with social isolation, and other stressors. This literature has been predominantly concerned with the possible mediation of long-term progression of coronary heart disease, and may not be quite so relevant as far as this acute triggering is concerned. When we study acute triggering, the same sorts of factors are possibly important, but in a rather different way, since short-term responses such as acute increases or decreases in activity, may be important.

I've tried to illustrate this possibility in this rather complicated cartoon. It shows at the top the emotional trigger, and at the bottom, the various clinical events we're interested in. These are preceded by a variety of pathophysiological effects such as plaque disruption, electrical instability of the heart which can lead to ventricular fibrillation, the formation of thrombus, and so forth. There are then a number of biological processes which might operate acutely in vulnerable patients to provoke these kinds of pathophysiological effects. This is really the challenge: to try and understand how to study these processes, and really tie down which factors might be important.

What sort of methods do we have for studying these phenomena? There are three broad methods of study which you will all be familiar with, because these are the primary integrative methods that we use in this field. One is to integrate biomarkers into epidemiological and population studies, which is very important, but which is not so well suited for this particular clinical issue because of the time course problem, and the fact that we might be looking at quite acute kinds of response. So we really have to fall back on two others methods: naturalistic monitoring -- that is, measurement of biological markers in everyday life -

- and also clinical experimental studies. I'd like to illustrate both of those methods in relation to emotional triggering.

Naturalistic monitoring studies are studies in which we look at the variation in levels of biomarkers as they relate to people's everyday experience. Instead of taking a single measure, such as a single blood sample, and using that in a large population, what we do is to take many measures in smaller study populations and look at covariation. The use of this method is really limited by technical issues having to do with what we can measure on a repeated basis in a relatively unobtrusive way. It is difficult, of course, to take repeated blood samples from people in their everyday lives and ask them to carry on as normal. It has been tried, but people really don't carry on as normal. So we are much more dependent on the noninvasive measures in which we're all interested here. The two which I'd like to focus on are heart rate variability and cortisol.

Heart rate variability is of great interest in this area because we know that impaired or reduced heart rate variability is a clinical risk marker, both for the long-term development of coronary heart disease, and also in people who have already had an acute cardiac event. People who show reduced heart-rate variability are at a much higher risk for future cardiac events and for poor prognosis. So this measure of the autonomic control of the heart is of considerable interest. Cortisol is related to many pathologies, and is relevant as far as cardiovascular disease is concerned in quite a variety of different ways. This is a summary that was made a couple of years ago, and shows you how cortisol seems to relate to many different pathological endpoints. So what we can ask in a naturalistic monitoring study is whether people who experience negative emotions – and in particular, anger, which is the one which has come out particularly in the emotional triggering studies -- show disturbed patterns of heart rate variability and cortisol output over the day.

This is the summary of a study that we've been doing recently. It's a relatively small study because this is quite an intensive type of investigation. As you can see, it involved 88 patients with suspected coronary artery disease. These were patients who were attending outpatient clinics. In England, we have things called Rapid Access Chest Pain Clinics, which patients are sent to by their family physicians if they have what look like early signs of angina or coronary heart disease. They were investigated, and had positive exercise tests or perfusion scans. We then studied the patients before they had undergone coronary angiography. This means that we didn't know whether they definitely had coronary artery disease or simply had chest pain, so were blind to diagnosis. What we did was to carry out 24-hour Holter monitoring -- that's 24-hour monitoring of the EKG and heart rate variability. W also took saliva samples for cortisol, and also measured people's mood over the day.

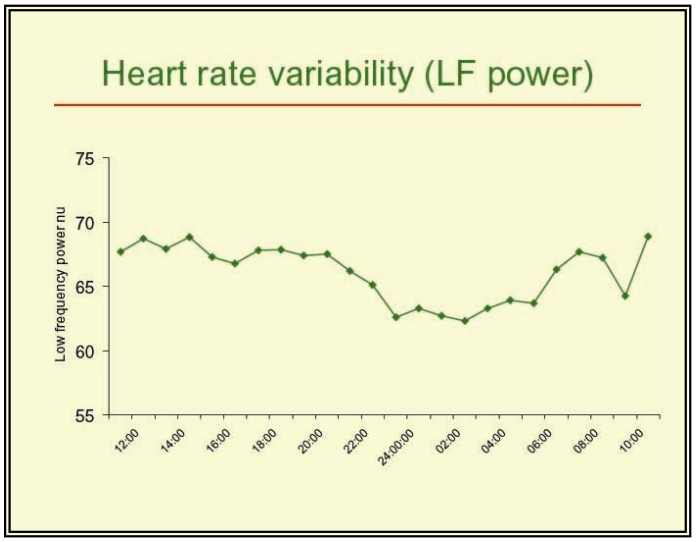
When one wants to look at mood over the day, various decisions have to be made about how to do that. Arthur Stone, who is here at the Workshop, has been the pioneer and champion of the use of ecological momentary assessments, which involve repeated measurements over the day, which are very useful in this context. But in this particular study, we wanted to use a method that was less obtrusive and had lower participant burden during the monitoring period. And so we decided to try out the Day Reconstruction Method, which some of you will also be familiar with, and which Arthur Stone was a co-author of the key paper published in *Science* in 2004.

"... more recently there's been a real expansion in study of other types of biological response – in particular, inflammatory responses. So there's been increased interest in looking at fibrinogen, CRP, IL-6, and heat shock proteins…"

I will describe the method briefly, because some people won't be so familiar with it. This is a method of trying to get an idea about people's mood and activities and social interactions over the day, using a retrospective recall technique rather than using measurements carried out actually during the period of interest. And the way we do it is to have an interview after the 24-hour monitoring period. The participant is asked to reconstruct their day, or the monitoring period, as a series of events, as if it was a film of their lives, and then to tell us about the different types of events which took place.

So here is an example. This is a single episode, actually the second episode in the morning for this person, which was timed between eight -ten, and eight-fifty in the morning. The person was commuting at that time. There are a whole lot of different activities they could

be doing, but these are the most common ones we've studied. The patients are also asked who they were with, where they were, and so forth.

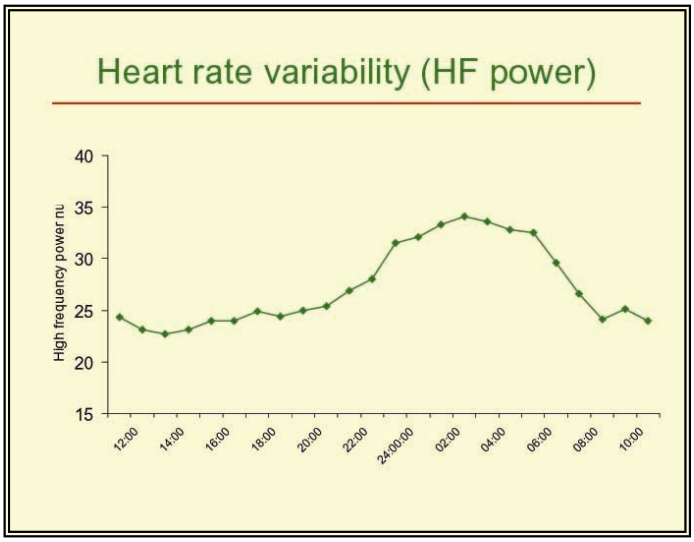


After that, we ask people to give us a series of ratings of what their mood was like during that period. These are the adjectives we used in this study, based on the original DRM – each was rated on 0 to 6 scale. This person said, "Well, for this episode, I was impatient for it to end. I wasn't very happy; rather frustrated. Wasn't feeling at all warm and friendly while commuting, and was moderately angry and hostile." With this sort of method, we can get build up an idea of how people are spending their time, and what they are feeling during that time. This can be analyzed a number of different ways, including looking at moment-to-moment changes. But I'm just going to show you some results based on looking at the aggregate levels of anger in these patients.

Heart rate variability is a measure for which the technology has improved enormously over recent years, and there are now a lot of methods for carrying out fairly rapid power spectrum analysis of records. Here is a record of a single patient from twelve o'clock one day until twelve o'clock the next day. What we've done here is to partition the power in the heart rate variability spectrum into different components, and then to produce this so-called waterfall plot. The particular components of interest are the high-frequency component between about 0.15 and 0.4 Hz, which is the component of the heart rate variability spectrum which is thought to be particularly related to parasympathetic, or vagal, activity. You can probably see these peaks here indicate that the parasympathetic activity is higher at night, because the sympathetic power is withdrawn at that point. Then there's low-frequency activity, which is in this area here – about 0.003 to 0.15Hz – which is thought to relate to sympathetic/parasympathetic balance. Then there's a very low component which is below about 0.003 Hz, which is often discarded in behavioral studies, although it actually contains a great deal of the power in the spectrum.

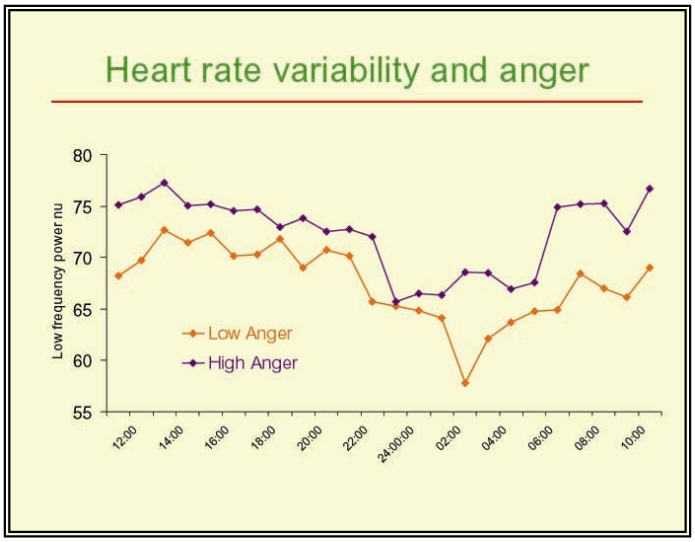
What does this look like over the day? These are data from these 88 people, and what I've done here is to produce averages over each one hour over this 24-hour period, or just under 24-hour period. This slide shows the low-frequency power, which is the component of the spectrum thought to be particularly related to the sympathetic/ parasympathetic balance, and so higher levels are associated with a greater sympathetic drive to the heart. What you can see is that low frequency power is reduced in the night compared with the day, and then, when people get up in the morning, it increases again. Whereas the high-frequency power shows the reverse effect; it's fairly modest during the day because the vagal tone is less dominant, but during the night there is this large increase in high-frequency power.

How does this relate to anger and other emotions? I'll just show you the anger results now. This slide has divided the sample of 88 people into low and high aggregate levels of anger, as measured with the Day



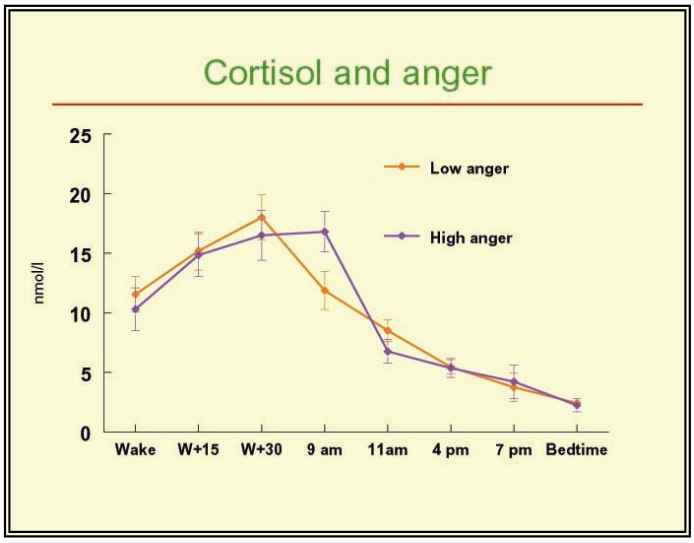
Reconstruction Method. So the orange line represents the low-anger people, and the high-anger people are in the purple color. What you can see here is that the low-frequency power, which is the sympathetic component, is higher all the time in the more angry individuals. So people who are experiencing greater

anger over the day have higher levels of low-frequency



power. In particular, levels don't go down as far in the night, and when people wake up in the morning -- this is around six am or so -- there is a large increase in low-frequency power. This is quite an important effect because acute coronary syndromes are particularly common in the early hours of the day. There's also an increase in the low-anger people, but it doesn't come up to such a high level.

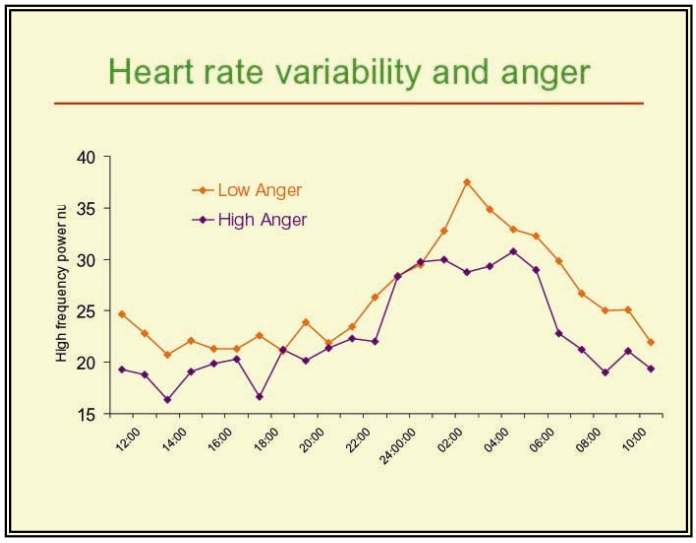
Now if we look at the high-frequency component of the spectrum, we essentially see the mirror image – that is, greater vagal tone throughout the time in the less angry individuals; a much higher level of vagal tone during the night; then it goes down when people wake up, but not to such low levels. So what this study suggests is that anger in patients with coronary artery disease is related to disturbances in this long-term



biomarker, and in rhythm disturbances, which are particularly relevant early in the day.

What happens with cortisol? This is the typical protocol we use for measuring cortisol, and there may be some discussion later in the day and tomorrow about what times are used. As many of you will know, there is particular interest in the waking response – that is, the change from waking over the first hour or so of the day. So we typically ask patients to measure saliva when they wake up, and then fifteen minutes, and then thirty minutes later, and then these are samples taken later in the day. This is a fairly typical diurnal profile of cortisol, which shows an increase after people wake up, and then lower levels going down to quite low levels at bedtime.

What we are particularly interested in, however, is this morning period here, which suggests -- and the

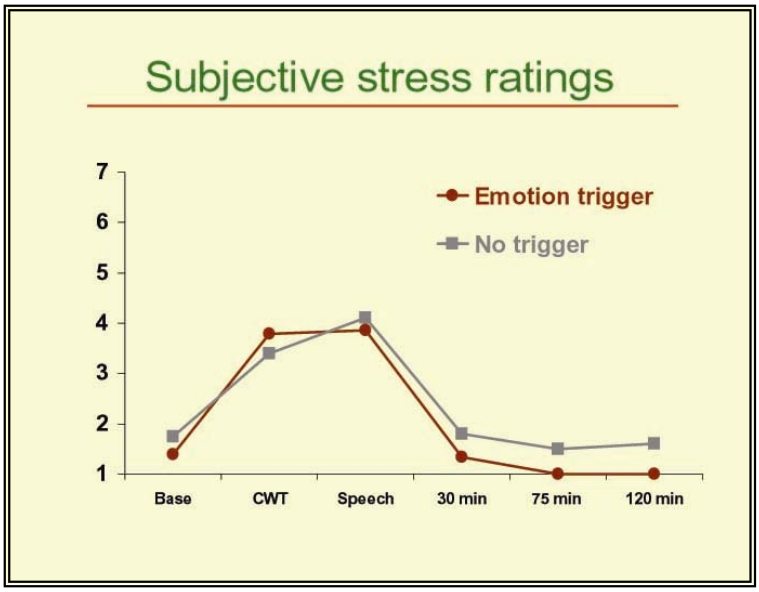


statistics do back it up -- that the angrier people show a less marked reduction in cortisol following the early waking response. The elevated cortisol seems to be sustained to a slightly greater extent in the angrier individuals. This is an effect that could be increasing vulnerability to acute cardiac events, particularly in this early period of the day. Later on in the day, the cortisols of high and low anger people go down together and run together fairly consistently.

So naturalistic monitoring studies can get us somewhere in trying to understand what the psychophysiological substrate is, not only in relation to coronary heart disease generally, but also in relation to acute cardiac events. But if one wants to really understand the acute responses, then we have to use more experimental types of methods.

We've been particularly interested in trying to integrate experimental techniques with population methods – nesting them in either clinical or population studies. We've been doing this particularly with the Whitehall II cohort, though I'm not going to talk about those data now. Instead, I'm going to talk about the use of experimental measures in this clinical situation of acute coronary disease.

We know that acute emotions elicit a whole range of biological responses which are relevant to coronary heart disease. The question here is whether these responses are related to a susceptibility to emotional triggering. Is it that those people who may be particularly vulnerable to emotional triggering show a different pattern of response? The particular measures we've been interested in are hemodynamic variables such as blood pressure, because of the sheer stress effect -- that's the effect on the coronary artery wall of an abrupt change in blood pressure – and also, platelet activation. Platelets are small white blood cells which immediately come together and cluster where there is any break in the vessel walls. So platelets are the first cells to come to any wound to try and stop bleeding, and they're also the first cells which come to any rupture of an intravascular plaque in order to try and plug that hole. Platelets can be in higher levels of activation, meaning that they're more likely to clump together. We are also interested in the inflammatory cytokines which are involved in these processes as well.

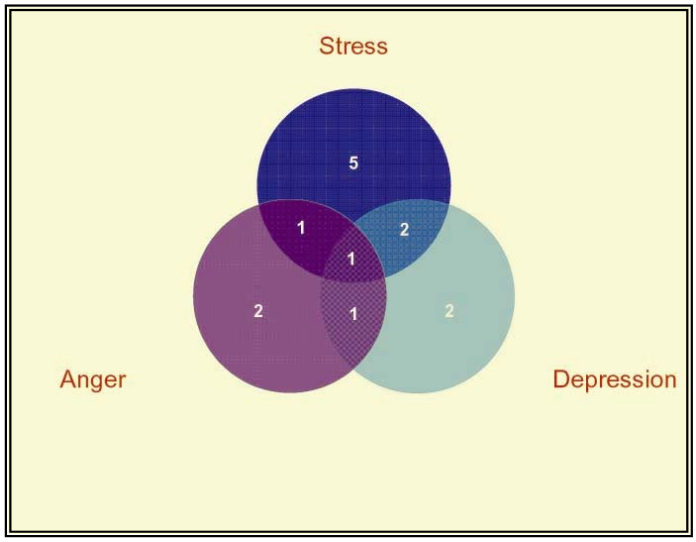


How does one study these variables in an experimental setting? It is unfortunately fairly intensive kind of work that is typically done with individual subjects, so these studies tend to take quite a long time
and can't be done very easily on large samples. But this slide shows you a typical pattern, just to give you an idea of the setup that is used. We have people coming in, take various physical measures, anthropometry and so forth, and then we usually put in a venous cannula and set up the other measuring equipment. Then the person is allowed to adapt for an extended period, so that the influence of that first procedure is past. Then there's a baseline period. Following that, people carry out various types of challenging tasks. Blood draws are taken at various points, and measures of subjective experience and other measures are done several times.

There's a lot of variation in this basic protocol. In particular, there's a lot of variation in the length of the post-stress recovery period. What we and others have found is that if one wants to study acute changes in cytokines such as IL-6 or tumor necrosis factor-alpha, then you really need to monitor people for quite extended periods because those peptides take quite a long time for biosynthesis and for increased levels to be observed in the circulation. Indeed, in our studies of cardiac patients, we usually have people continuing for two hours following the termination of tasks. This is a long period for people to sit there – three to four hours altogether – and that's quite a labor-intensive kind of activity.

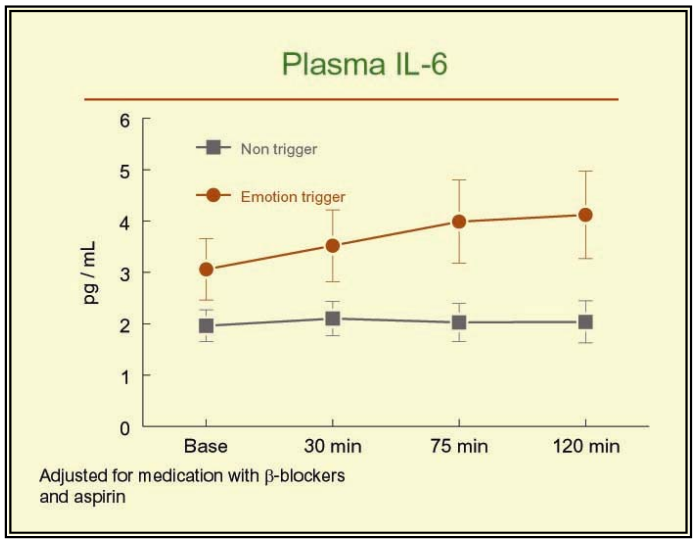
What are the biological responses to this procedure? This slide summarizes a study we carried out a couple of years ago, comparing patients with coronary artery disease and controls. This shows the blood pressure responses in patients and controls. The particular challenges that were included here were mirror-tracing, which is a psychomotor task, and a simulated public speaking task, where people were given a scenario and they were asked to generate an impromptu speech which was to last for three minutes. Some people find this quite stressful. And they are told that it's being filmed, and that their performance will later be rated for its competence and fluency. Typically we use a scenario involving someone being falsely accused of shoplifting. They are just leaving a big department store, and the store detective arrests them and says that they've just stolen some knickers from the ladies' lingerie department, and they have to go to the manager's office, and then they have to vindicate themselves in three minutes. People find this quite challenging. As you can see, in both controls and patients with coronary artery disease, blood pressure goes up quite a

bit, but the patients with coronary artery disease show



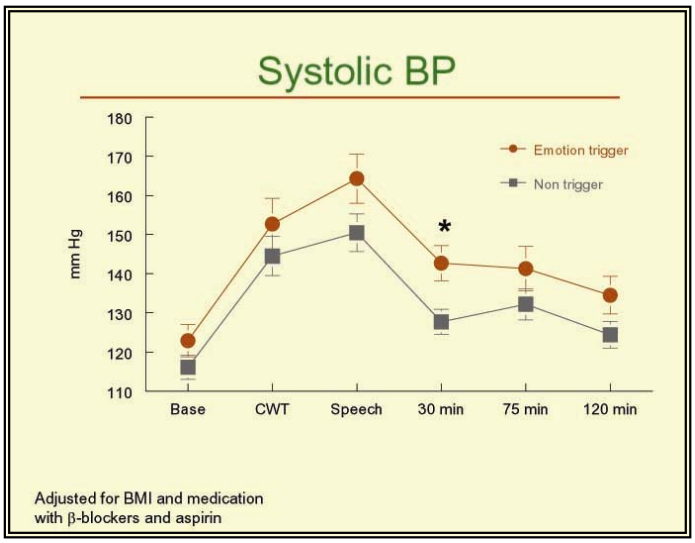
larger responses.

In another recent study, we harnessed this method to look at the issue of emotional triggering. What we did, essentially, was to compare patients who had or hadn't reported an emotional trigger prior to their acute coronary syndrome. These patients were all from the earlier clinical study that I described. What we did was to test 34 male patients about 12 to 18 months after discharge from hospital. They were certainly recovered from their acute coronary event. We then administered an experimental stress protocol. If you're doing these procedures with clinical patients, there are a number of ethical issues to consider. You have to do this type of work rather carefully in the presence of a physician with all the appropriate safeguards in place, because it's possible that the sorts of biological responses you're provoking could produce a new cardiac event, though,



thankfully, that has never happened to us. Of these 34 patients, we had fourteen who reported different kinds of triggering experience in the two hours prior to their acute coronary syndrome. So the question we asked was whether they would show any differences in their psychophysiological response compared with the rest of the population.

Let's look, first of all, at subjective stress ratings. We always take these measures to find out what people are actually experiencing. This slide shows a simple rating of stress on a 1 to 7 point scale. You can see that these tasks provoke a moderate increase in the subjective experience of stress, but with no difference between groups in the magnitude of response. So the people who had previously experienced emotional triggering are not just more emotionally labile as far as their subjective experience is concerned.



There were, however, important differences at the psychophysiological level. One of these was in systolic blood pressure. What you can see here are slightly higher levels of systolic blood pressure throughout the procedure in the trigger group. And in particular, in the period of the immediate post-challenge recovery period, the reduction is slower in the trigger people. So the provocation of the psychophysiological response seems to be more extended in those individuals compared with controls.

A particularly striking difference was seen in the activation of platelets. As I said earlier, platelets are involved in the early stages of acute coronary syndromes. What we found in this study, using a method of assessing platelet activation with flow cytometry, is that there was an acute increase in

activation of platelets, but only in the emotional trigger patients. So even though all these patients had experienced a major coronary event, the non-trigger group did not show a change in platelet activation level at all. The emotional trigger group showed more or less a doubling of platelet activation level, not only in blood taken immediately following the stress tasks, but even thirty minutes after that. That suggests that they are psychophysiologically prone to respond to emotion with a different sort of profile which could be one of the factors which puts them at increased risk of acute coronary events.

I will show you one more result from this study, which is not a published result, concerning plasma IL-6. As I said earlier, changes in IL-6 tend to take a fair time to evolve. So this slide shows the mean levels of IL-6 at baseline, then thirty minutes following stress, and then two hours following stress. What you can see is that there is an increase in the emotional trigger group but not in the non-trigger group. The extent to which these changes in plasma level reflect the speed of active involvement of IL-6 in inflammatory processes is uncertain. It's likely that IL-6 at the local level may respond more rapidly. We have some data from gene expression studies suggesting that that responses can occur within thirty minutes of stress. But what these data suggest is that an inflammatory milieu is provoked by emotional triggers, but particularly in the patients who appear to be sensitive to emotional stimuli.

Well, let me just close with some brief remarks about whether these responses matter as far as our understanding of clinical problems is concerned. I think they matter in three ways. First of all, we want to understand better the etiology, epidemiology and psychobiology of coronary heart disease. In a way, this literature on acute triggering has co-existed with the long-term literature on psychosocial factors, and we don't really know how those two aspects link together. We know, for example, that social isolation is associated with reduced heart rate variability. Are those same individuals also vulnerable to acute triggers? So we need to bring together the longer-term etiological literature with this more acute triggering literature.

Secondly, I think this type of work is relevant to patients' quality of life and future adaptation. We know from our own studies that people who experience acute emotional triggers show somewhat worse adaptation as far as quality of life is concerned in the six to twelve months post-infarction. So this is definitely an important issue as far as patient care is concerned.

Then, of course, there is the question of prevention, which is not something I'm going to go into now, though we can certainly discuss it if you're interested. But there are ways in which people have begun to think about the extent to which this information about triggering can be used to improve prevention. It's a major problem, because even though emotional triggers are present, acute coronary syndromes are still fairly rare events, and so thinking about large-scale prevention is quite a challenge.

Let me close with some broad conclusions. As I've tried to explain, anger and various kinds of acute stress seem to be able to act as short-term triggers of acute coronary syndromes. This work complements the longer-term literature on psychobiological and psychosocial factors and coronary heart disease. I would also argue that we can use psychophysiological studies to understand the underlying processes. We can use both the naturalistic studies in everyday life to look at such things as disturbed autonomic control and neuroendocrine function, and we can use more intensive, experimental studies to look at whether there are distinctive biological stress response profiles in people who are at risk.

I will finish by acknowledging various coworkers. Thank you very much.

[*Applause.*]

Q&A With Dr. Steptoe

Laibson: There are two extreme models that might apply to your findings. One extreme model let's call the bulb blooming in the spring. And you can imagine the very hot day in April that causes the bulbs to bloom. And all that that's doing is accelerating the bloom by a few weeks. It would happen in late April. Instead, it's now happening in early April because of this extremely hot day. The other model would be the light bulb breaking model. If I jostle a light bulb, I may break it, and if I just somehow did not jostle the light bulb on that day, it's as good as a new light bulb. We think it's a constant hazard of failure.

Now, it makes all the difference which of these two models is relevant in the cardio-vascular disease case, and I was intrigued by the evidence on the earthquake

in Los Angeles because it looked as though the heightened frequency of heart attacks in the period around the earthquake was actually followed by a period of depressed frequency of heart attacks, which would suggest that it's more like the spring bulb than the light bulb. But I think you'll probably be able to tell us how you think about the two extremes and where we lie on the spectrum.

Steptoe: Thank you. That is a very important issue, and one I didn't mention. But as you say, what one really needs for this, the only way you can deal with this problem is to use a population-based approach. It's very difficult to do it with the more individual level kind of research. As you pointed out, in the earthquake case, that is the best evidence for a kind of brought-forward response, the fact that those people were already vulnerable and that they might have had an acute event over the next two or three weeks, but it happens to have been brought-forward.

And that's a model that's been used, and particularly, looking at life events and depression, for example, or life events and schizophrenia breakdowns, that kind of brought-forward model. We don't actually know, because there aren't enough population studies, and it's really necessary, I think, to try and study acute cardiac events at the population level in order to understand which of these models is really operating.

Now the psychophysiological literature suggests that there may be certain people who are vulnerable, in which case that would tend to suggest that it's a characteristic of individuals, and that is not necessarily going to operate just in terms of short two to three week acceleration of effects. But, this is speculation at the moment. So this is certainly an area in which we need to combine methodologies.

Hendrie: I was just wondering how you translate this to epidemiological studies, and one way would be to measure the susceptibility to emotional triggers. And I was wondering, Dr. Steptoe, what measurements do you use to get that susceptibility measurement?

Steptoe: There isn't, at the moment, a good measure of that. Basically what one is doing in the clinical setting is asking people about their experience during these acute phases of illness. And you could have someone, for example, who says that they were very angry in the two hours prior to the onset of their current cardiac syndrome, but they were also very angry the day before, and on the day before that, because they're just angry people. In this case, there's no reason why that acute episode should be a trigger at that particular moment. So we can't just use the subjective report as an indication of susceptibility in this respect.

Hendrie: Just to follow up. I'm very interested in measurements, but there used to be these old anger prone measurements that looked at personality traits. Is that something similar?

Steptoe: Well, we do have measures of trait anger or anger proneness, and people who score high on these measures are likely to report anger more often in response to relatively neutral stimuli, so this could increase risk. That's certainly true.

Harrison: I was wondering if you could comment more broadly about the issue of perturbations in population studies, because it seems that there are opportunities that people can take advantage of that are stressful within their own studies. For instance, in the Age Gene Environment Susceptibility Study, we decided that we would try to measure heart rate variability as people went through our standard examination, which included a 45-minute cognitive test, and a fifteen- to twenty-minute physical function testing of strength and physical performance. And we don't have the data yet, unfortunately.

But I think that there are opportunities like this that people can take advantage of. But in general, it seems that it's harder to standardize perturbations in population studies, and yet the responses to perturbations may be, in fact, much more interesting than just background data because I saw that with your cortisol levels. It looked like the only difference, really, was that 9:00 a.m. cortisol, and otherwise the mean cortisols, both before, at nadir, and at zenith were pretty much the same. So I was wondering if you could just comment generally about the issue of perturbation in population studies.

Steptoe: Yes, I think that's something we've been struggling with, along with many other people. The strategy you mention, which is actually to take measurements during the other types of assessments that are being carried out, is one which I think is quite a promising procedure. For example, in aging studies, one is often doing neurocognitive types of tests, of cognitive function, which many older people find quite

challenging, and it's certainly possible to use these as a perturbation, as an acute perturbation in the system. The problem is whether you have biological measures which are really going to be sensitive to those types of change.

Heart rate variability is probably a measure which one could use in this context, but something like cortisol level is more difficult. We've measured cortisol levels before and after interviews and physical function and cognitive tests, and memory tests in older people, and find very small changes compared with the responses that you get in a psychophysiological testing in general. So unless one has very large samples, I think it's quite a problem to see how well that's going to work.

Harrison: Could you try measuring things other than cortisol like platelets or anything else?

Steptoe: Unfortunately, I should mention that in platelet measurements, we use have two problems. First of all, they require blood samples, and secondly, they require laboratory flow cytometry within two hours, so have to be measured very acutely. This is unlike measures like cytokines, where we can spin plasma down and freeze it; platelets have to be assessed at the time, and this really does limit use in larger scale studies.

Q: My model of how this happens might be that of my favorite literary masterpiece, "Alexander and the Terrible, Horrible, No-Good, Very Bad Day." So my question is whether you've looked at the joint occurrence of events, overeating and stress, this, that, and the other thing, rather than looking for a single indicator of what brings on MI?

Steptoe: Yes, I think that is a very important issue. We haven't looked at combinations, unfortunately, but we're just starting a new study, a much larger study, in which I hope that we'll be able to do this because it surely must be an important component. For all these acute events, things like time of day and day of the week are also very important.

Q: Thanks very much.

Steptoe: Thank you very much.

[*Applause.*]

Panel Discussion: ***Global Approaches to Integration of Biological and Psychosocial Measures from Population-Based Studies of Aging***

Stacy Lindau, MD, MAPP (University of Chicago)

Introducing the National Social Life, Health and Aging Project:

Biological Measures and Approaches to Integration of Biological and Survey Data

This workshop grew out of a need for the National Social Life, Health and Aging Project. In 2002, we were in the early phases of planning and had made the decision to incorporate biological measures into that study. There was general enthusiasm among our group for doing that, but that we also needed some outside expertise before we went into the field. We had to start by making sure we were speaking the same language. When Ed Laumann suggested we incorporate "biomarkers," I, as a physician, thought he was talking about tumor markers. As a gynecologist, a biomarker to me was something we use to monitor a woman's treatment response to ovarian cancer, for example.

It took me a little while to understand what my social science colleagues meant by biomarkers. What it meant was collecting the kinds of biological data that we frequently obtain in the clinical setting, but in the home setting or in combination with collecting detailed survey questionnaire information.

So we invited a very small group of people -- I think there were twenty of us all together, including those of us involved with the project -- and we talked about the measures we were considering, what were the gold standards, what things we had not considered, etc. That meeting was a very important part of the crystallization of the biological component of the NSHAP study.

We're now five years later, and this conference has grown from a focused consultation to the NSHAP study to one where we all can learn from each other's studies. In these five years, I haven't really been a formally presented on the National Social Life, Health and Aging Project. Of course, I'm only one of many people, many of whom are in this room, who brought this project to fruition, and it's my honor to be able to tell you a little bit about it today.

Because the time is short, what I will do is give you some background information, as others have just done about their studies, and then tell you what I am seeing in these data from my viewpoint, from the perspective of someone interested in advancing older women's health.

Before I venture into that, I want to acknowledge the members of the Biomarker Core, the staff members who have made this conference possible. We have a wonderful group of young people – many of whom are students on a path to medicine and social science careers. In my day-to-day life, this group is what makes the Biomarker Core a worthy and rewarding venture.

"...it meant collecting the kinds of biological data that we frequently obtain in the clinical setting, but in the home setting or in combination with collecting detailed survey questionnaire information."

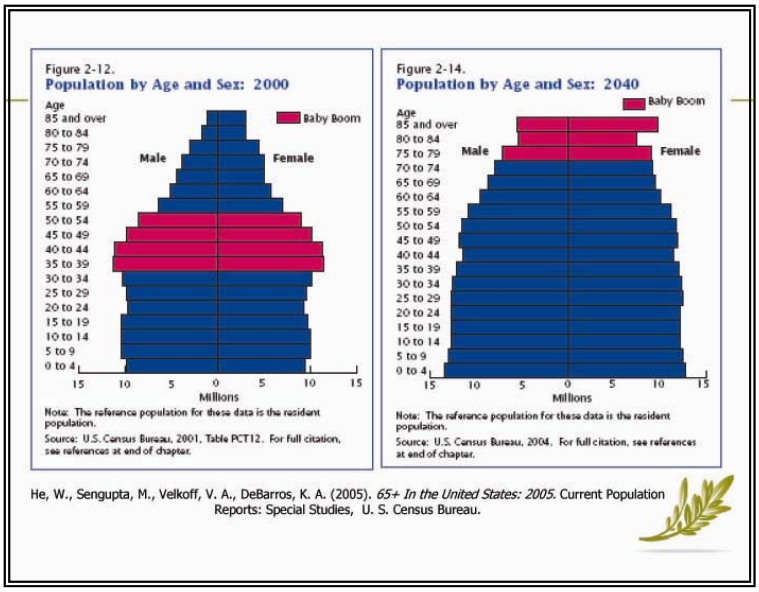
Now, I will introduce you to the National Social Life, Health and Aging Project and present just one project that incorporates both the biological and the social measures we've collected in NSHAP as kind of a case study. If there's time, I have put together some thoughts about the theoretical and historical foundations for what we call "integrated health research." I've done that because, in reviewing past year's Proceedings, and particularly last year, I realize that we weren't necessarily using the same words to mean the same things. So if there's time for that, I'll include it. If there's not, then maybe through comments, or if somebody else doesn't show up for their talk, I have one in the bag…

To pique your interest in the rest of the talk, I point out that the NSHAP Wave I data will soon be publicly available for your use via the National Archive of Computerized Data on Aging (NACDA). I want to acknowledge Jim McNally, who's here with NACDA and is a close contact of NSHAP as we prepare to

make the data publicly available. We're very excited about this and hope you'll find some elements of relevance to your research.

I want to acknowledge the NSHAP collaborators. Linda Waite is the study's Principal Investigator. And the many others you see listed here. Unfortunately, there isn't a slide big enough to begin to list all of those who've been involved with the project.

We also, to accomplish the collection of a very broad variety of biological measures, had to affiliate with a number of investigators and laboratories. Many of these investigators have become collaborators on the research, which has been terrific. One of the downsides of incorporating biological measure collection in a large, population-based study is having to coordinate the logistics of samples and specimens going all over the place. One of the upsides, alternatively, is this outreach has connected us to individuals who have expertise and resources that we would not otherwise find locally. One example of this is our collaboration with Thom McDade, and the growing relationship between the Biomarker Core and C2S at Northwestern.



This also took a good deal of creativity and resources in combination with those of our primary funder, which has been National Institutes of Health, particularly the National Institute on Aging. Corporate and individual contributions and grants, which we acknowledge here, have been important to our success.

We were officially funded in October, 2003. We conducted a pretest in 2004. We learned a great deal from that pretest, which helped us implement a very successful project. Our field period spanned from June, 2005 to March, 2006. Stephen Smith and his team at NORC and a team of close to 150 interviewers, were extremely successful in recruiting 3,005 men and women for the study, with a 75.5% weighted response rate.

Our analysis phase began in October, 2006. Now that we have a data set ready for public release, we are in the process of reanalyzing to make sure that what we find will be the same as what you find when you work with our data. It's now that we're really entering the trajectory of scientific productivity with the data set.

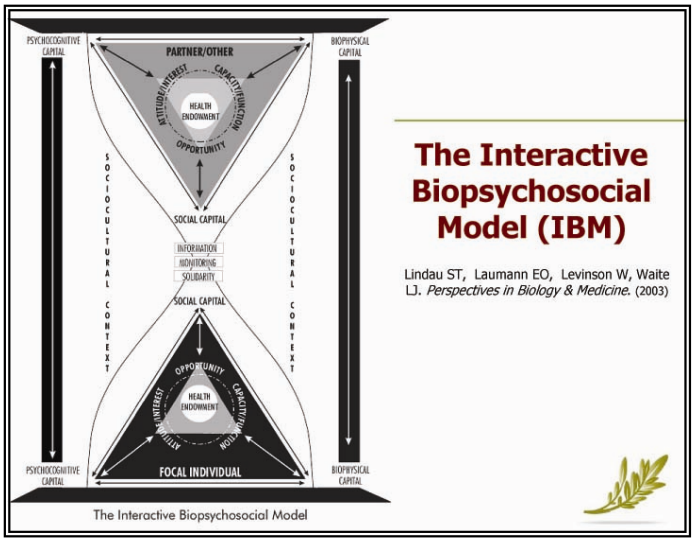
So why did we do a study like NSHAP? Well, there are many reasons. I'm particularly interested in issues pertaining to older women's health, and many of you are demographers. You're all familiar, obviously, with the unprecedented shift in the age structure of the United States population (also occurring in Western Europe) and the disproportionate number of women as compared to men in the oldest age groups. I'm interested in what the implications of this are for public health and for women's health. This is the demographic context for this work.

There are a number of conceptual models that inform the work we are doing. A tie that binds is the question of what are the biological mechanisms through which social factors, any one of many, influence health, health outcomes, and trajectories of health and disease. We came together, a group of social scientists and physicians, thinking about what data we would need to collect to start to get at those questions, and we found that we needed a common vocabulary to move forward.

We built from existing models. We talked about the "medical model," that came out of 16th century rationalism, (and was coined the "medical model" in the late 20th century by people who offered alternative models). George Engel offered the biopsychosocial model (1977), a commonly-referenced model that informs medical practice today.

The medical model is described in current literature as a model that's based on a biological basis of disease, so a derangement in some biological entity, system, or network of systems results in a disease or an illness, and the story starts and ends there. George Engel, who was a psychiatrist trying to figure out where

psychiatry fit into the larger body of medicine in the late 1970s, argued that this model was restrictive. The biopsychosocial model, alternatively, says we ought to incorporate factors beyond biology, including social



and psychological factors, in understanding the root causes of disease.

What he didn't talk so much about was the reverse of that: What are the *effects* of disease on not just the biological system, but also psychological and social factors? And there wasn't much detail in terms of what he meant by the social factors in particular.

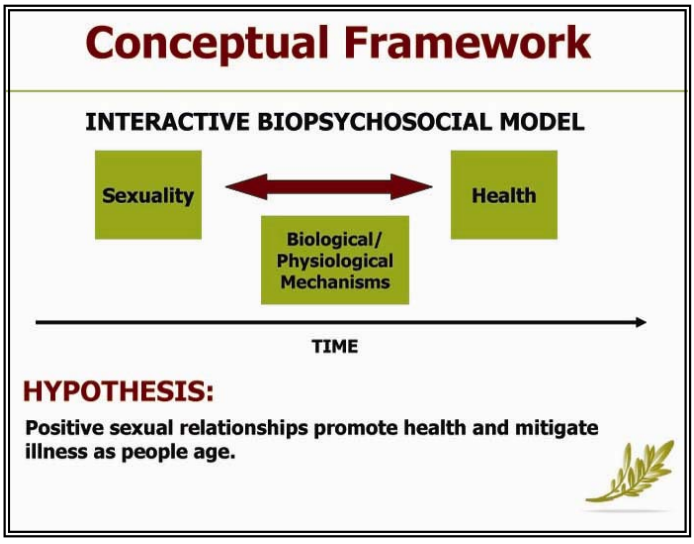
In building on that work and work of others (Teresa Seeman's, for example), we developed a model to frame out work, which we call the Interactive Biopsychosocial Model. The National Social Life, Health and Aging Project aims to decipher how social relationships influence health at older ages. What was missing from prior frameworks was the explicit recognition of an important social other in conceptualizing health.

In all the models I mentioned to you before, the focus is the individual -- an individual patient, really -- who seeks medical treatment or who has a problem. In this model, we think about health as jointly produced between two individuals. Health is conceived positively, working from the current World Health Organization definition. Because of an interest in studying sexuality at older ages, this is between an individual and his or her sexual partner, in most cases the spouse. But the important social other doesn't have to be the sexual partner. A paper we published a few years ago, at the start of this project, describes this model.

The study had originally been designed to look at dyads, as the model would suggest. Of course, we need to collect data from both partners in order to understand how health is jointly produced. The realities of funding resulted in a study where we only interviewed one member of the dyad. Information about the partner was obtained by reports from the focal respondent. My colleagues, as they go on with this study, will look to include the social other, if possible, in Wave 2.

This gives a frame of reference in terms of the thinking that underlies the NSHAP study. It also helps explain why we took somewhat of a departure in the biological measures we collected. There are a number of commonalities in the kinds of biological measures collected in other studies described here today. Many of these studies are informed by models relating to allostatic load or physiologic dysregulation theories. Pathophysiology would be a somewhat synonymous medical term. The measures obtained are guided by these perspectives and often invoke "stress" as a common variable of interest.

Our study wasn't primarily designed to measure these types of biomarkers, although we do collect many measures that might fit into models where those theoretical perspectives apply. Instead, we collected



biological measures in the domains of sensory function and in the domains of physical function related to sexuality; for example, vaginal mucosal characteristics and salivary sex hormone measures.

As I mentioned, the Interactive Biopsychosocial Model

is our framework for understanding how important social relationships and health are related. We hypothesize that this is a bi-directional or reciprocal relationship. I emphasize sexual relationships in the

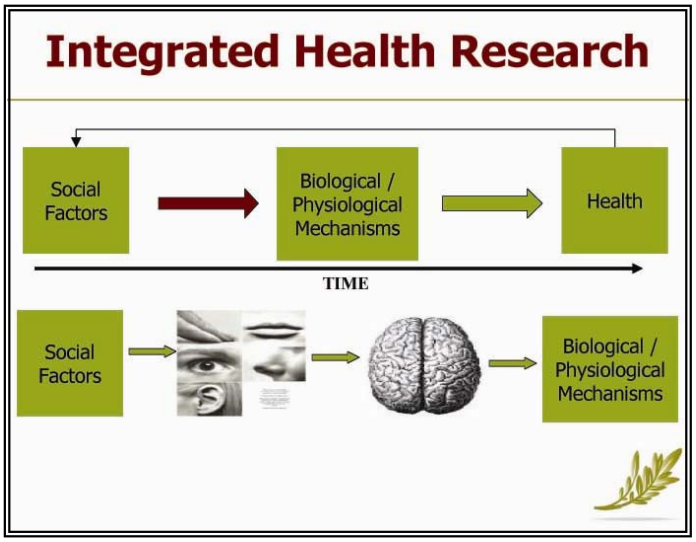


diagram because that's my particular area of interest. And what are the biological mechanisms that we measured? We're going to tell you that in just a moment.

Our longitudinal hypothesis is that positive sexual relationships promote health and mitigate illness as people age.

So, what is integrated health research?

In the case of NSHAP, we're interested in how social factors are perceived by individuals, which is why we went to great lengths to measure sensory function. We use measures of olfaction, taste, vision, touch. And we tried to get an objective measure of hearing, although time ran out before technology could catch up with us, and used a validated questionnaire measure to assess hearing. We didn't directly measure how the brain works, although exciting clinical research using fMRI technology is doing this. We look upstream at sensory perception. This is then processed in the brain and results in the downstream biological processes that we can measure. So until we figure out how to get a functional MRI machine into our wheelie bags.

The study design, as I mentioned, included over 3,000 community-residing adults. The people were derived by a population-based probability sample. We oversampled African American and Hispanic individuals, and the interview was offered in Spanish for those who preferred that. The in-home interview lasted about 120-minutes. We collected the biological measures in the home, and we left a questionnaire behind for individuals to complete themselves. We had about an 80% response rate on the leave-behind questionnaire.

The estimated population distributions are adjusted for differential nonresponse and differential participation. These estimates are comparable to U.S. Census data in all the categories. Self-rated health is that standard "excellent, very good, good, fair, poor" question. How good is your health? And the distribution for the self-rated health reflects that seen, for example, in the Health and Retirement Study.

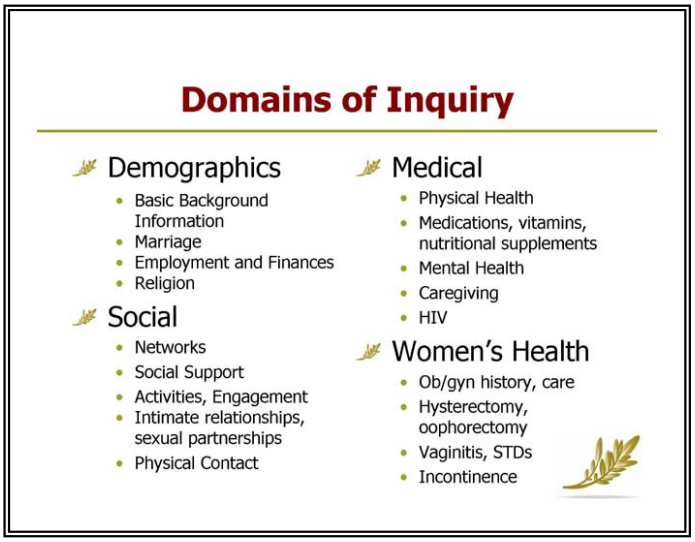
The broad domains of inquiry are shown here. But this gives you some sense of the categories of information that we collected in the study. We have good detail in many of these areas. We have uniquely good detail in the area of social networks. We obtained comprehensive detail on sexuality and sexual relationships. Again, this is somewhat unprecedented in a population-based study of older adults in the United States. We also obtained a lot of information about health -- the directly-observed medication log, in particular, is another strength. And because I'm interested in women's health, I highlight some of the areas where we ask questions pertinent to older women's health.

Est. Pop. Distributions (%)

	Men (n=1455)	Women (n=1550)
AGE		
57-64	43.6	39.2
65-74	35.0	34.8
75-85	21.4	26.0
RACE/ETHNICITY		
White	80.6	80.3
African-American	9.2	10.7
Latino	7.0	6.7
Other	3.2	2.2
RELATIONSHIP STATUS		
Married	77.9	55.5
Other intimate relationship	7.4	5.5
No relationship	14.7	39.0
SELF-RATED HEALTH		
Poor/Fair	25.5	24.2
Good	27.5	31.5
Very good/Excellent	47.0	44.3

So what specific biomeasures did we collect? These are categorized by the specimen rather than by the physiology. There are a number of ways to slice and dice the rationale for collecting certain measures. We collected blood, saliva, and vaginal swabs. Vaginal self-

swabs have been used in younger populations for STD - (sexually transmitted disease) testing, but it was a novel method to use with older women in the home. Anthropometrics, physiological measures, sensory measures, as I mentioned, gait and balance comprised



the physical measures obtained in NSHAP. I want to mention salivary cotinine just in case any of you are not familiar. This is an important metabolite of tobacco and a biomarker of tobacco use and exposure. Here, you see the cooperation rates for the biological measures. The first column identifies those who were eligible. Some of these measures were modularized in a random way so as to allow us more time in the home. We didn't have enough time to do everything, and so we randomized some of these measures to subsets of our group. This explains why some of the numbers eligible are smaller than others.

You can see that the cooperation rates overall were excellent. I largely attribute this to two things: one, a tremendous amount of work developing smooth protocols; and a lot of investment on behalf of the team, especially NORC, in finding and training the right interviewers.

Incidentally, we found -- this is not a scientific observation, but at least a strong anecdotal one that I think most of us agree about — that medically trained individuals such as nurses or doctors on hiatus were not the best interviewers. Lay people who are good interviewers, who know how to get in the door, who know how to establish trust and rapport, really anecdotally seemed to be those who had the skills we needed to be successful with the biological measure collection. And I'm sure Katie Lundeen, Angie Jaszczak or somebody else, if they have a different view, would comment when I finish.

We had relatively lower response rates on the more invasive tests. We offered HIV testing using a transmucosal exudate specimen, with 89.2% willing to be tested for HIV. As expected, the HIV positivity rate was very low in our sample.

Blood spots -- 85% allowed us to use a finger stick mechanism for collecting blood that was then placed on filter paper. That protocol was developed in conjunction with Thom McDade and Sharon Williams. For the vaginal swabs -- 67.6% of women provided a vaginal specimen, most of which were satisfactory for our assay procedures.

Another factor that influenced how we chose the biomeasures we did was whether or not we could collect them in a minimally invasive way. I was very staunchly dedicated to this, because I feel that the only way we're going to make progress in expanding the number of biological measures that we can collect with lay interviewers in remote areas is if we continue to push ourselves to do it with alternatives to venipuncture -- for example, with blood spots, with urine-based measures, saliva, etc.

And so Thom and I have done some work together, building on earlier work in defining the principles of minimal invasiveness with Jenna Mahay. These criteria

NSHAP Biomeasures Cooperation

Measure	Eligible Respondents	Cooperating Respondents	Cooperation Rate*
Height	2,977	2,930	98.6%
Weight	2,977	2,927	98.4%
Blood pressure	3,004	2,950	98.4%
Touch	1,502	1,474	98.4%
Smell	3,004	2,943	98.3%
Waist circumference	3,004	2,916	97.2%
Distance vision	1,505	1,441	96.0%
Taste	3,004	2,867	95.9%
Get up and go	1,485	1,377	93.6%
Saliva	3,004	2,721	90.8%
Oral fluid for HIV test	972	865	89.2%
Blood spots	2,493	2,105	85.0%
Vaginal swabs	1,550	1,028	67.6%

*Person-level weights are adjusted for non-response by age and urbanicity.

guide my decisions about what kinds of biological measures I'm willing to entertain in the data collection design.

So this framework helps me think about -- more from a strategic sense -- why we collected the biomeasures

that we did, and why do it in a population-based

Applying Biomeasures in NSHAP

Uses of Biomeasures	Population-Based Sample	Clinic-Based Sample
To detect and monitor risk for disease, pre-disease, disease, mortality OR to quantify and monitor function	++	++
To recruit or exclude people from study	-	++
To determine *efficacy* of intervention	--	++
To determine *effectiveness* of intervention	++	+
To identify biological correlates or mechanisms of social/environmental conditions	++	--

++ = Very well suited -- = Poorly suited

sample versus in the clinic where it's much easier to get biological data from people. And obviously, this audience is aware of what the strengths and weaknesses are of collecting data in the clinic versus in the population, and Dr. Steptoe's talk highlighted that.

In our case, to detect and monitor risk for disease, pre-disease, mortality, or to quantify and monitor function, a population-based sample works really well. You can do this in a clinic-based sample, but there are obviously important biases and the generalizability is really fairly limited to other people accessing care in similar clinic settings.

There are several ways of thinking about biomeasures or biomarkers and their importance. Especially when you come from the perspective of somebody

Disciplinary Use of Biomeasures

Uses of Biomeasures	Population-Based Sample	Clinic-Based Sample
To detect and monitor risk for disease, pre-disease, disease, mortality OR to quantify and monitor function	RISK: Genital HPV EPIDEMIOLOGY Sex hormone metabolism	BIOMEDICINE
To recruit or exclude people from study	-	
To determine *efficacy* of intervention	--	
To determine *effectiveness* of intervention	EPIDEMIOLOGY vaccine, new hearing devices	
To identify biological correlates or mechanisms of social/environmental conditions	BIODEMOGRAPHY/ANTHROPOLOGY Restricted Mobility	

++ = Very well suited -- = Poorly suited

conducting a trial for pharmaceuticals, biomarkers might be used to recruit or exclude people from such a study; to determine the efficacy of an intervention, meaning how a drug works under ideal circumstances. That's a more appropriate domain for a clinical study, and that's not something NSHAP attempted to do.

On the other hand, biomeasures can be used to determine the effectiveness of an intervention -- say, a new vaccine. We collected biomeasures pre- and post introduction of the vaccine into the population to determine if it prevents infection.. In this case, a population-based sample might be really useful. And then finally, what most of us are doing is collecting biomeasures to identify correlates or mechanisms of social and environmental conditions.

Some examples from NSHAP in the first category

Applying Biomeasures in NSHAP

Uses of Biomeasures	Population-Based Sample	Clinic-Based Sample
To detect and monitor risk for disease, pre-disease, disease, mortality OR to quantify and monitor function	RISK: Genital HPV Tobacco use Obesity FUNCTION: Mucosal integrity Sex hormone metabolism	++
To recruit or exclude people from study	-	++
To determine *efficacy* of intervention	--	++
To determine *effectiveness* of intervention	Future public health interventions: e.g. smoking cessation, HPV vaccine, new hearing devices	+
To identify biological correlates or mechanisms of social/environmental conditions	Hypertension Impaired Glucose Metabolism Restricted Mobility	--

++ = Very well suited -- = Poorly suited

would be looking at measures or markers of risk. So HPV, tobacco use, obesity -- these are risk markers that we can ascertain from the data we collected. We can also quantify aspects of physiological function -- for example, vaginal mucosal integrity or sex hormone metabolism.

If there were a major public health intervention targeted at smoking cessation for older adults, an HPV vaccine, a new hearing device that was going to be paid for by Medicare and widely distributed, we may be able to determine to some degree the effectiveness of those interventions if we were in the field collecting relevant biological data before and after. And so you get the idea.

The point I want to make here is that all of our studies,

and NSHAP fits this description, are engaged in multiple scientific activities that touch on the traditions of epidemiology, of clinical biomedicine, of biodemography and anthropology. And whether we have one of these experts on our team, I think it's important to recognize that these are the areas where we're venturing; most of our studies are working in these domains.

In terms of NSHAP, you see a picture of our wheelie bag here. We found a very efficient way to put all the materials in one bag. We created this "laboratory without walls," which included the home, and then involved a diaspora of specimens sent to geographically dispersed clinical laboratories for analysis.

"These criteria, for me, guide my decisions about what kinds of biological measures I'm willing to entertain in the data collection design."

To give you a brief glimpse into how these data are being used, let's look at our study pertaining to cervical cancer risk among older women human papillomavirus is regarded as the most important cause of cervical cancer. Cervical cancer is a very big problem in the world outside of the United States and outside of Western medicine. It's not a leading causes of cancer death in women in the United States, because we have screening. Few women die. Older women in the US are less likely to have cervical cancer but disproportionately more likely to die from it, as compared with younger women.

Using the NSHAP data, we can look at vaginal presence of human papillomavirus from the vaginal swabs. We couldn't look biologically at the outcomes of HPV, like dysplasia or cervical cancer. For these, we rely on self-report, which is limited as a measure of those two outcomes.

There are a number of biological factors that we thought were important in terms of determining whether we would find HPV in older women. There were a number of social and behavior factors, primarily tobacco use and exposure, which we thought we ought to understand.

I would propose that what you see here is an integrated model for understanding whether or not an older woman has HPV. In the clinical study, we do a test. She has it or not. We treat her accordingly. The population setting and interdisciplinary collaboration allows us to think more broadly about variables that might be influencing HPV infection and opportunities for intervention.

In order to collect vaginal swabs in the home, we had to adapt methodologies that had been described for younger women. Selection of the specific swab and the directions we gave women had to be tailored to the particular assays of interest to our study. We worked with an illustrator, Rachel Seelen, who helped us develop tailored instructions that a participant would bring with her to the bathroom to help collect the specimens.

Once we received the specimens, and many of you can relate to this, so now you've accomplished the hard task, right? Getting the specimen into a cold pack and into a mailbox and into one laboratory, and then from that laboratory to the next laboratory was also part of the challenge. And fortunately, we were successful with that. In this case, with HPV, the swab landed in the laboratory of microbiologist Jeanne Jordan. She's been a wonderful collaborator. Jeanne is the one who did the HPV analyses.

So how do you detect HPV once you've gotten an older women in her home to give you a swab and get it to the lab? Well, initially, we were using one test called the Hybrid Capture II Test, which told us whether there were these cancer-causing types of HPV or not. Since the study began, a new assay has become available to us. It gives us more detailed information. It gives us a better understanding of not only whether HPV is there, but what kinds of HPV are there.

And that raises questions and challenges. In this case, on the specimens that tested positive with the clinically-approved test, we did decide to also test them using this new pre-clinical technology.

I want to mention, in response to Gerry MacQuillan's earlier comment, that NHANES is receptive to helping investigators as they design their studies. She's been a terrific resource for us. And we've been able to collaborate, such that the data we collected are comparable with NHANES, which make them all the more useful to the rest of the research community.

In addition to looking for HPV in those specimens, we

also wanted to look at a direct measure of the function of the vaginal epithelium. There are different cell types in the vaginal epithelium. In order to understand those, we needed to get those swabs to a lab at the University of Chicago -- Martha McClintock's lab -- where we could put the cells onto slides, get them stained in a clinical laboratory — and then code them to asses the degree to which the mucosa was atrophic or very thin and poorly estrogenized, or not atrophic, or well-estrogenized. And you see the spectrum here. These are beautiful cytology slides from NSHAP participants.

I have to say, I was optimistic about getting women to provide us with the specimens. I was optimistic we could get them where they needed to go. I was really not sure what would happen by the time they went through all those steps. We're very, very pleased that we have, really, the equivalent of clinical specimens here. Not all of them worked out. We definitely had some attrition. But we have a lot of really great data that have never before been collected in a population setting.

Once we obtained the cells for analysis, we needed to code them. We did this qualitatively by eye, looking at the different colors of the cells and measuring the relative proportions of cells. Another way we do this is quantitatively by using software that allow us to more rapidly assign codes to the slides. The technology is evolving not just in the assay we're using, but also the possibilities for coding techniques.

"The biopsychosocial model, alternatively, says we ought to incorporate factors beyond biology, including social and psychological factors, in understanding the root causes of disease."

So we find HPV. We now can explore the cytological correlates of HPV. And we have several sex hormone correlates; they're the biological correlates. We also want to know, are there any behavioral correlates of HPV? And then we have cotinine, the measure of tobacco exposure, which is a behavioral correlate. So we have to process all of those data make them useful for an integrated analysis.

In looking at the salivary sex hormones (this is very much the work of Natalia Gavrilova, who has helped us explore these), we first have to look at the distributions, determine outliers -- all the things you're familiar with -- and then figure out whether we ought to be thinking about each of these hormones separately, or whether we ought to create integrated scores of endocrine function. I must admit, we haven't really figured that one out yet, but we're working on it.

"...but to some degree, the markers we're using are the markers other people have used, and that's why we keep using them."

With regard to cotinine, the behavioral marker, tobacco is an independent risk factor for Human Papilloma Virus infection, and we think that it also acts synergistically at the cellular level to promote HPV activity at the level of the vaginal mucosa. We're interested in knowing whether there's a correlation between the objective measure of smoking behavior and HPV infection.

We had to do a lot of work with the help of Elizabeth Gaumer, who is a doctoral candidate in sociology, and Mindy Drumm, a biostatistician at the University of Chicago, just to figure out what the salivary cotinine data meant. And ultimately, in combining self-reported smoking behavior with the salivary cotinine behavior, we now have a picture of the distribution of cotinine. The next step is to figure out how to integrate the cotinine with the self-report with HPV.

Finally, we're interested in the relationship between inflammation and immunity with HPV infection. Now, why do we have CRP and Epstein-Barr Virus (EBV) antibody titers as our markers of inflammation and immune function? Are these the best, most powerful markers of inflammation and immune function? Maybe, but they're the markers that were available for use with blood spot analysis. So I admit that openly, because as a physician, I'm sensitive to the fact that CRP is just one marker on a cascade of probably hundreds that indicate inflammatory response and inflammatory function.

Not to negate its importance and its increasing relevance with regard to, say, cardiovascular disease or in clinical use, but to some degree, the markers we're using are the markers other people have used, and that's why we keep using them.

Finally, we have several self-report measures to integrate. Those include the demographics, the social and sexuality variables, and a variety of health measures.

So I want to just give one example, to which I've been alluding, of the challenge of integrating data -- in this case, self-reported smoking and the cotinine measure. What we find is that there is a strong correlation between cotinine and self-reported smoking. This is good. This makes us feel like the cotinine marker works.

"...we're interested in how social factors are perceived by individuals, which is why we went to great lengths to measure sensory function."

We also found there's a significant correlation in univariate and in multivariate analysis between self-reported smoking, current smoking, and HPV. But we find, at least so far, that there doesn't appear to be a significant association between the biomarker for smoking and the presence of high risk HPV. Why is that, what does that mean, what do we do about that? If you know the answer, you're another person I want to meet while I'm here.

So with regard to just this one small example from a vast array of, I think, very exciting data from the NSHAP study, there are some next steps that we need to take in order to understand predictors of this oncogenic, very prevalent virus, about which we know almost nothing in older women. And these are some of those steps.

There are some challenges to integrated health research summarized here, from my perspective. These are ones that I have been thinking about, and others have written about. So we have a wonderful interdisciplinary team, we spent several years together collecting data across the spectrum: social, behavioral, biological. We finally have the data, and the tendency -- and I'm a victim of this as well -- is to approach the analysis in a much more segregated way.

Integrating ourselves together, solving problems at an interdisciplinary level, is a lot harder than getting the variables you know well, publishing them in the journals where you have a track record, etc. And I think there's nobody better than the investigators of the team that collected the cross-disciplinary data to be able to jointly analyze and make use of those, and yet we're fighting the forces against that.

The other is -- and I become increasingly aware of this as I'm reviewing articles and proposals -- a notion of expertise across disciplines versus expert collaborator. What I hope to become, which takes a tremendous amount of work, is a quality collaborator with my social science colleagues. I tried to get a PhD; the institutional forces worked against me in doing that. There are a few of us who cross disciplines with dual degrees MD's and PhD's or PhD's in biology and PhD's in social science, but those people are few and far between.

Developing ourselves and developing our students as expert collaborators across disciplines is an important goal for success in making use of these cross-disciplinary data. It's a non-trivial challenge to find the right group of people to review cross-disciplinary studies and give meaningful feedback or integrated analysis.

"Integrating ourselves together, solving problems at an interdisciplinary level, is a lot harder than getting the variables you know well, publishing them in the journals where you have a track record, etc."

These are tremendous challenges. On the one hand, we have a great opportunity in collecting these biological data. On the other hand, there's a lot of work we'd have to do in order to make these biological data reach their maximum potential in terms of improving public health. The moving target of technology in light of institutional barriers, presents an ongoing dilemma. So I will end there. I welcome any questions, and thank you very much.

Q & A with Dr. Lindau

Waite: Any other questions? People are hungry for lunch. Okay, so thank you very much for giving me an opportunity to tell you about the study. Please, if you're interested in working with the data, they will be available soon, as I mentioned. I, Linda Waite, several of the individuals sitting in the back of the room, in the very last row, are very familiar with the data set. Colm O'Muircheartaigh – I'm not sure if he's left – are

all individuals you can approach if you're interested in working with the data.

Waite: And we can certainly send you a PDF of the questionnaire and a [leave behind] and everything so you can be perusing that in the next couple of weeks while all of the data are ready [to go].

Lindau: Also, NORC is working on a web site where I believe these materials ultimately will be accessible and through NACDA as well. Not just the data, but also information that will help you work with the data. Phil Schumm is not here today, but he's done a tremendous amount of work as a leader in making that data set useful for you. I saw another question. Yes?

Q: Do you have a recommendation for the analysis of the data in the various modules? So if I understand the leave behind questionnaire correctly, there are modules that are given to selected subjects. So is your recommendation to just use the cases that have the complete data, or are you, as a team, developing some kind of multiple imputation strategy to complete the data for those who purposefully didn't get the questions?

Lindau: I think we're somewhere between the two. And I wish Phil were here to answer the technical question. But the modularization happened in two ways. It affected the biomeasures that were collected, and for those individuals we have no backup. They either were modularized or weren't. And imputation could be done, for somebody who had the analytic skills to do it. We are not currently putting out recommendations for imputation.

With regard to the questions that were modularized, an appendix to the data set clearly talks about which questions were modularized, which ones were in the core questionnaire, so you'll know that information. And for our use, we have been using the modularized questions, but thinking about adjusting for mode effects. Linda, did you have something to add to that?

Waite: We decided to get certain measures only for enough people so we could describe the population distribution. So we didn't measure vision on everybody, and I don't think there's really a way to… we did get self-reports in all those cases, but I don't think there's a way to really say what measured vision would have been for the people. I suppose you could predict it, given that you do have self-reported vision and measured vision for those 1,000 people.

But then, in addition, we modularized some questions to be randomly assigned either to the in person or to a leave behind. And there we do expect, and I think people have to be sensitive to, mode effects. So some questions could be…the answers could differ somewhat, depending on how they were to ask it. And so I would encourage people to take that into consideration, at least including an indicator variable for mode of collection for those questions in their analyses.

Lindau: Well, thank you, everybody.

We're going to make our way over to lunch, which starts at noon. Lunch is in Room 621, so up to the sixth floor. And then Hanna, what time will the walk occur? Is it 12:45 p.m.? Twelve-forty. Hannah is the leader. And so anyone who would like to have some non-sedentariness in their day is welcome to join us for a walk to Millennium Park.

Short Presentations Part I: ***Novel Methods in Home-Based Biomarker Collection***

Arthur Stone, PhD (Stony Brook University)

Well, thank you. It's a pleasure being here today. I'm just trying to look at what time I should stop. Just so I get it right.

Alright. Well, I'm from the Medical School. Sorry -- I have to do these disclosures or else the people at the Medical School are unhappy. I'll go back if anyone wants to see them again. What I'm going to do today is tell you a bit about our experiences in measuring patient-reported outcomes -- self-report -- using a conceptual scheme that actually was developed not only for self-report in the real world, in a person's natural environment, but it was also developed for use in measuring physiological outcomes as well. And what I'm going to try to do is share some of the lessons that we've learned from self-report research in the field to the collection of -- what are they called -- biomeasures now.

So three questions in this brief presentation: Why might we care about being in a natural environment? What sorts of biomeasures are available for natural environment assessment? And what kinds of methodologies are available for biomarker sampling in the environment.

So why do we care about being in the natural environment in the first place? Well, I have to say, it probably only matters to you folks if the biomarker that you're looking at is changing or has the potential to change somewhat rapidly. If we're talking about a biomarker that changes very, very slowly -- say, over the course of a year -- then some of what I'm going to say probably isn't very relevant. The focus today is going to be on markers or phenomena that change rather rapidly. I'll be talking about changes within a day, but I think the lessons are scalable to within a week or to within a month. You'll see.

So why do we care about this? Because biomarkers may change according to social circumstances, the kind of things that are going on in a person's life, the physical environment, where they are, what they're doing, other characteristics of the environment. And they may change according to time: time of day, day of week, etc. These are all considerations.

Now, what we really want are self-reports, and our biomarkers are to be ecologically valid. This goes back to Brunswick and his concept of representative design. And we want to make sure that when we are taking a biomarker at some point in time, that it is representative of what we want a generalization of whatever time span we target.

And we know that this can be a problem. Take the measurement of blood pressure. You all know about the phenomenon of white coat hypertension, where you get very different answers if you measure people in the real world than you do with a casual arresting blood pressure in a laboratory. You may not know that there's also something called masked hypertension, which we're studying at Stony Brook and is the opposite of white coat hypertension. But the point is, you want to avoid these kinds of effects in your measurement of biomarkers.

"Why might we care about being in a natural environment? What sorts of biomeasures are available for natural environment assessment?"

Another thing that we want to be able to do very often is assess real-life environmental and psychosocial influences. So, for example, we may be very interested in dietary influences, social interactions, activity settings, medications, stress -- the measurement of cortisol, for example -- very often these are the kinds of potential influences that we're interested in.

Another reason that we may want to be in the real world is that we want to know something about the diurnal variation of biomarkers. Almost every physiological process has a rhythm of some sort. The field of chronobiology is concerned with such associations.

And we also might care about variability in our measures. It might not be important what the level of a measure is (for example, it's average), but it might be important how variable it is, as in heart-rate variability, for example. And this is something that we probably should assess in the field under natural conditions, because it's just difficult to do in the laboratory.

Now, what kinds of biomarkers are available? Well, this is just a brief sampling of the kinds that I thought

about in preparing this talk: blood pressure, heart rate, actigraphy, skin conductance, all sorts of assays that are available from saliva, glucose levels from newly developed sensors that you can wear and that take rather rapid measurements of glucose levels.

There's a relatively new item called the "life shirt," which is a complex device that one wears, and it monitors something like 200 physiological parameters based respiration. And here are some examples of how nifty these things are. That little device towards the top measures both heart rate and actigraphy – activity levels. Andrew Steptoe uses something like this in his studies on a one-day basis. We use these to record for up to fourteen consecutive days.

Now, what I'm going to focus on today is methodologies for sampling from the real environment. And I think that a lot of times in biomarker research, we don't think enough about this. So the concept or framework I'll use is the one that I mentioned at the outset, which was developed primarily for self-report. It also applies to physiological monitoring, and that's ecological momentary assessment – EMA. Not ESM; that's Experience Sampling.

"Another thing that we want to be able to do very often is assess real-life environmental and psychosocial influences."

EMA is based on multiple samples of people in their natural environments. And one of the main parts of EMA is very, very careful consideration of how you're going to do the sampling: so that you don't get into trouble with the design, and you accomplish your goals. In the field of self-report, EMA, which does real-time assessments of self-report, is terribly important because we're trying to avoid retrospective bias when we have people think back and try to summarize over a period of time. That's not really the case for the phsyiological measures. Nevertheless, the issue of ecological validity is still very, very relevant.

The other thing with real-time monitoring that's important is, in addition to taking the physiological measurements, you need to take record information about the things that might be affecting the physiological measurement. So if you believe, for example, that recent smoking or recent eating or exercise is going to affect your physiological variable, then you may want to have a self-report associated with taking that sample so that you can look at that and figure out to what degree is the level or variability or whatever it is you're looking at of your biomarker being influenced by those activities.

If you want to know more about EMA -- this is my unabashed plug -- the analysis of that kind of data, this book came out about a month ago -- called *The Science of Real-Time Data Capture*, published by Oxford. It's a very nice overview.

Now, there are many forms of real-time data capture. You can use diaries, you can use interactive voice response, the Internet, cell phones. Another system -- and I don't know if it's going to be talked about here – is telemedicine: for example, the kind of systems that the VA is using which incorporate home-dwelling devices that actually then are linked via Bluetooth networks to other kinds of sensors in the environment that can be very useful. And that may be a topic that's worth considering.

But let me tell you a little bit about use of an electronic diary as the technology to drive the EMA sampling, which will include for the biomarkers, as well. So to the right you see a Palm-type device. You could use a cell phone, by the way. And basically, it contains a program that will administer a series of audio prompts which are telling a person, "Look at me, look at this device. I'm going to tell you what to do: what to tell me about, what questions to answer, and whether or not I want you to take a physiological sample or not." So you can imagine linkages of this with cortisol, with taking a saliva sample, or you could activate a blood pressure monitor, or you could do all manner of collections.

By the way, there is a new development in the field. There are several companies that make this kind of devices. They are used by the pharmaceutical industry quite a lot, which is part of the reason for that disclosure. There are now devices that are linked via Bluetooth and direct lines so that you have the EMA device that's driving the sampling connected directly to the physiological monitoring device, and there's two-way communication so that you can actually get people to tell you where they are and what they're doing at the time that their physiological index, and that's very, very neat, and creates some possibilities of designs. But you're going to see a bunch of designs in a minute.

Many of you are probably not familiar with EMA. Let me just give you a brief overview. And I give you an idea of some of the screens that are getting at what kinds of activities are you engaged in at the time that you were beeped. We'll get to how you get beeped in a minute. And you can see: you might be working, doing housework, overseeing household services, on the telephone, sleeping, grooming – a whole variety of things.

The point is that you can record a lot about the social environment, about the physical environment, about whatever it is that you're interested in. And these are what are called check boxes. You can check as many of those as you need. You can do miniature VAS-type scales, which are very useful for pain and for affect. You can do what are called buttons, just to give you an example.

So let's move now to the sampling protocols. You can't just think about a sampling protocol without thinking about the reason for doing the study in the first place. So there are three main reasons for wanting to do an EMA biomarker-type study. One is that you're trying to characterize a person's level in general. So you're interested in, for that week or for that day, a person's overall cortisol level. Let's call it Type 1.

You might also be interested in characterizing the variability or the slope over time. That's a second kind of measure that you might want.

Third, you might be interested in within-day associations, where you're interested in, for example, how does a particular activity or environmental characteristic go along with a biomarker. That's within-person association. The kind of sampling you do, the kind of design you put together, is dependent upon which of these three kinds of questions or combinations of questions you're interested in.

As I already mentioned, we're going to be focusing on day as the unit of measurement. But the principle can be generalized to over the course of a week, over the course of a month. I'm not going to get into some of the practical issues, but if you want, there is a MacArthur network website that applies some of these principles to the measurement of cortisol: how many samples do you need per day in order to get at levels, how many samples do you need per day and when to get at diurnal cycle, etc. These are issues that are very relevant for people who are doing HPA access research.

There are three kinds of samples that one can do: *interval-contingent*, which is based on a time sampling, simple time. So I'll give you some examples of these, but I just want to give you an overview first. Possible problems with this is that people might anticipate, if you were doing this over time for many, many days, when the samples were going to be taken, and they may change their behavior.

"...methodologies for sampling from the real environment. And I think that a lot of times in biomarker research, we don't think enough about this."

Event-contingent: This is very useful because what you do is you ask a person to take a sample according to whether or not an event has occurred. Examples of events are whether you've had an argument with a spouse, whether you're commuting, whether you've just exercised, etc. Some issues here are that it can be difficult to confirm compliance with the schedule, because you really don't know whether the thing happened or not. You're basing it entirely on the self-report.

Third is *signal-contingent*, which is what I've sort of been talking about where you get a beep, and the beep or a vibration or whatever it happens to be signals you to take the biological measure.

Now, let me give you some examples. One kind of sampling is end-of-day sampling. End-of-day is nice because you can do this without too much burden on a person for many, many, many days. It's been used in the world of pain to record people's daily pain or affect over the course of even up to a year. You can see that there are issues. Let's say that you took one sample; you were interested in blood pressure and characterizing the day. Would taking a sample at the end of the day be an ideal way to characterize blood pressure? Probably not, for obvious reasons. That's why you need to start thinking about other designs.

Here's an example of interval-contingent. Let's say that we were interested in characterizing cortisol levels throughout the day or cortisol levels, and we were going to take it at eight a.m. and eight p.m. every day.

Would this be a reasonable way of getting that information? Would there be enough samples? Would the samples be spaced at a proper interval? Would you be able to characterize the slope, for example? Probably not.

Here's an example of an event-contingent kind of design. Let's say that you were interested in hormone levels right after a migraine attack. You might have a person self-report a migraine attack using one of these Palm devices, and then take a saliva sample. Now, issues with that, then, is, you'll get hormonal levels right after the migraine, but you won't know much else about what hormonal levels were like. And that's where we're going to end up. We're going to end up with more complicated designs that combine different kinds of sampling.

Here's another example of an event-contingent, looking at urinary incontinence and hormone assessment, if that happened to be what you were interested in. Signal-contingent: This is closest to the experience-sampling method. This is where you randomly choose when you're going to sample a person throughout the day. The program is set up so that it does whatever number of assessments you want at random intervals throughout the day.

"The point is that you can record a lot about the social environment, about the physical environment, about whatever it is that you're interested in."

There's a version of this which is called stratified random sampling, and that is that you demand that the program take random samples within intervals so that you make sure you're covering the entire day. That's very, very useful if you're interested in diurnal patterns. But you have a random component in there so that you have full generalizability; that you are sampling over situations and the person isn't picking when they are reporting.

Combined modalities are very, very neat because what you can start to do is miniature within subject type experiments. Let me just take you through this very quickly. Here is an example where we're interested in a hormonal assay associated with when a person is craving for cigarettes. Let's say our theory was that hormones are changing and they are associated with craving. So the design here is a person reports when they have a very strong craving for a cigarette. That's the purple "E." And you might have a saliva sample taken immediately after that.

But, then, what do you compare it to? There's no "within" design component yet that's of interest. So you combine this with a random sampling of saliva samples and what you can then do is a comparison between the random samples and the event-driven samples, those that are associated with craving, and start to put together a pretty good picture about whether hormones are different or the same around the time of craving.

This one also is kind of an interesting design. It's a variant of event-driven and interval-contingent. Let's say that you were interested in whether or not different kinds of events or different kinds of people had a different response to a large event -- say, an argument with a spouse. And you hypothesized that there would be a differential decline in the level of the hormone you were looking at according to the kind of person -- let's say, males or females.

So you might have an event reported and then have the computer do an interval-contingent -- let's say every twenty minutes or so -- signaling of the individual in order to collect saliva samples so that for everyone, you would have the event and then you would have "occur" following the event. We've done this using saliva and looking at cortisol, actually.

So what I've tried to do in a short amount of time is give you some sense of how the principles of EMA can be useful for biomarker research. It's certainly very, very feasible. The sampling framework is incredibly important. It deserves a lot of thought. One should imagine what the data's going to be looking like when you're done with the study; what comparisons you'll be making. And EMA, I think, offers a way to determine what kind of sampling you might want to do.

So I think I'm just on time. Thanks very much.

[*Applause.*]

Q&A with Dr. Stone

Q: You said that this is also used for dietary histories?

Stone: I collaborate with a group at the University of Pittsburgh that's doing dietary monitoring. And there, the issue is the self-reporting retrospection. People are not very good at remembering what they've eaten in detail over more than one day or so. So the food diary business is now moving to a 24-hour recall period, and some people are using Palms to do that.

Q: Can you say anything about how using this device affects the behavior of the individual?

Stone: Yes. By the way, let me be sure you understand something that perhaps was not clear in my presentation. The electronic diary was just one example of many, many different ways that you can signal people in the real world. I just want to be clear about that.

Q: Yeah. So how generalized…Like this device or devices like it?

Stone: Thank you. Now I'll answer the question.

Q: Okay, thank you.

Stone: The main considerations here are other reactive arrangements that are set up. In other words, is doing this thing over and over again going to change the phenomenon that we're measuring? There are four studies of reactivity in the area of pain. Three of them were kind of secondary analyses of data, trying to figure that out. One study that I published in the journal *Pain* a few years ago was specially designed to look at that issue. And so far, there's no evidence of a large reactivity effect. That's in direct conflict with what's known as the self-monitoring literature, which uses self-monitoring as a therapeutic tool.

And although no one has studied this, there is speculation that at least results are very different how you set the study up. If you tell people, "I want you to be self-monitoring and paying attention," and I'm self-monitoring something that we're working on therapeutically, I think that pulls for reactive arrangements. If you're not doing that, I think it much less pulls for reactivity. But that's speculation. But the data to-date suggests that there is not a large effect, at least in the area of pain and chronic illness.

Q: Thank you.

Stone: Yes?

Q: This is the input data. So what are some of the things you mentioned? You mentioned saliva, but what are some of the other physiologic measures that can be triggered by something like this? You know, sort of the measurement side of it. Because this is the reminder side, but, then, what are some of the things that you've plugged into this or that other people have plugged into this?

Stone: Well, we're running right now…in the area of masked hypertension at Stony Brook, we're running a thousand people through a one-day protocol where they're wearing ambulatory blood pressure monitors and they're doing EMA at the same time. In another study, we are monitoring actigraphy, cortisol and heart rate in a three-day study where we're doing the same kind of thing.

Pretty much anything that a person can do on his/her own can be used in this framework.

Tamara Harris, MD, MS (NIA)

Well, I can honestly say that I don't have anything quite as neatly packaged that I'm going to be presenting. And I'm also not going to be talking about collecting energy expenditure measurement in the home because I've never done that. And I nominated the best person I could think of who has done something close to that, who is Amy Luke, who is going to speak next.

So, I thought instead what I would talk about are some issues regarding biomarkers in studies of aging. The first thing I wanted to say is that at this conference, although everyone here is interested in biomarkers, I really would like to say that I don't think every study needs to have biological specimens collected in it in order for it to be a valid study.

And I think that the collection of biomarkers, especially if there's no specific hypothesis is associated with it, is expensive, time-consuming, adds a tremendous level of complexity, as Dr. Lindau has talked about this morning, and there are a lot of analytic headaches. There's a question of what do you do with your repository when your study is over. And I think that we have to remember that there are many questions that can be well investigated without detailed biologic measures or specimens, and that there seems to be sort of a cyclical emphasis on biologic variables, especially because of the emphasis now in genetics, genomics and proteomics.

And it's not clear to me when this cycle will break, but I think there is a lot of research that doesn't need this. And I think the other thing that we need to point out is that these specimens are all gathered in context. They're useful for many things, but they're not necessarily useful for all things, and they are affected by the condition of the patients and also by the circumstances under which they're collected.

Biological measures - Setting

Measure	Mail/phone	Home	Clinic	Alerts
Biologic specimens	X	X	X	X
Anthropometry	X	X	X	X
Blood pressure, AAI	--	X	X	X
Electrocardiogram	--	X	X	X
Spirometry	--	X	X	X
Performance measures	--	X	X	--
Strength	--	X	X	--
Dental	--	X	X	X
DXA of heel	--	X	X	X
DXA of body or hip	--	--	X*	X
Magnetic resonance imaging	--	--	X*	X
X-ray of hands, hips	--	X	X*	X
X-ray tomography	--	--	X*	X
Retinal photos	--	--	X*	X

So, why include biomarkers? Your hypothesis requires it. That's really the most important reason. You want to remove the effect of the examiner or the examination because people feel that the physiologic measurements tend to be more unaffected by the self-report; or alternatively, you want to capture the effect of the examiner and the examination, which is another reason, as Dr. Steptoe talked about this morning. People regard these measures as being closer to physiologic function so that they can increase the specificity of classification.

And I think that sometimes, especially if people are not experienced in the use of biomarkers, they tend to attribute properties to these measurements that are not true of self-report. And in fact, there are just as many problems and issues with the collection of biomarkers as they are in the collection of self-report. They help with determining the severity of impairments or conditions, and they tend to be value-added in studies. But they're not always better than a report.

So one example that I can give that has been a huge controversy in geriatrics is the question of whether you can use reported function as an outcome or whether you have to do performance measures. So the reported function might be, can you walk across a room versus timing people while they do a four-meter walk. And it turns out that actually, both measurements give information, they don't completely overlap, and that one doesn't actually represent exactly the same universe of outcome.

And performance measures have several issues that self-report don't have. For instance, they're voluntary. People have to get up and they have to walk. Some people walk more quickly; some people walk more slowly. Cognition plays a role in performance measures. The action may be unfamiliar. There are places in the world where people don't use chairs. So asking people to do a chair rise just doesn't really make any sense. And there are also practice effects that

come into the data after you look at performance times.

In addition, when choosing a biomarker, either performance tests or laboratory tests, they really have the same issue in terms of trying to identify what marker is going to capture the population that you're looking at. So for fit elderly people, using the four-meter walk or six-meter walk really doesn't have enough ceiling effect so that everybody really is at the maximum, whereas if you have a frail population, you can't use the 400-meter fast walk because nobody will be able to do it.

"And I think that we have to remember that there are many questions that can be well investigated without detailed biologic measures or specimens..."

And the same thing happens when we look at laboratory tests. When we're looking at a lot of laboratory tests, the clinical assay suffices, which is that essentially, people have very high levels of whatever it is that you're looking to measure, and the assay can be relatively insensitive and will still give you information. And that was true of the earlier measurements of C-reactor protein. But now we know that there's a whole world of epidemiology that couldn't be even approached until we had high sensitivity assays, because we were really looking for variation in the normal range. So I think in addition to Dr. Stone's discussion about thinking about the question, we have to think about who the population is in terms of designing the measurement.

Then the other thing I wanted to mention -- I didn't have time to put it on another slide -- is the issue of double-dipping in measurements. And I do think that it is possible to, for instance, as Dr. Stone mentioned, you can put one of the heart measurements on people. You get the benefit of both the movement as well as the heart rate variability. And I think that things like that can be very useful in terms of adding extra value with respect to population studies.

The other thing I want to say is you can also send instruments to the home. And probably the most intensive measurement that I've seen people do is to actually conduct sleep studies. The sleep, heart, health study and the Honolulu aging study both conducted polysomnography in the home. And if you can do polysomnography in the home, you can really do almost anything.

So, because this is a very invasive test, it means wiring people up and hooking them up while they're sleeping at night. The other thing about biomarkers is that the more passive the biomarker is, the more likely you are to have success with doing these measurements. And when I say "passive," what I mean is that you're basically instrumenting the person at rest.

So these kinds of things would be specimen collection, electrocardiogram, anthropo-metry. There's no real level of cooperation the person has to do. They don't have any effort at all. A dental exam, a retinal photograph, the heart rate variability, the activity monitoring, endothelial dysfunction -- all of these things can be done to people passively. And in fact, there is relatively little missing data for the passive measurements.

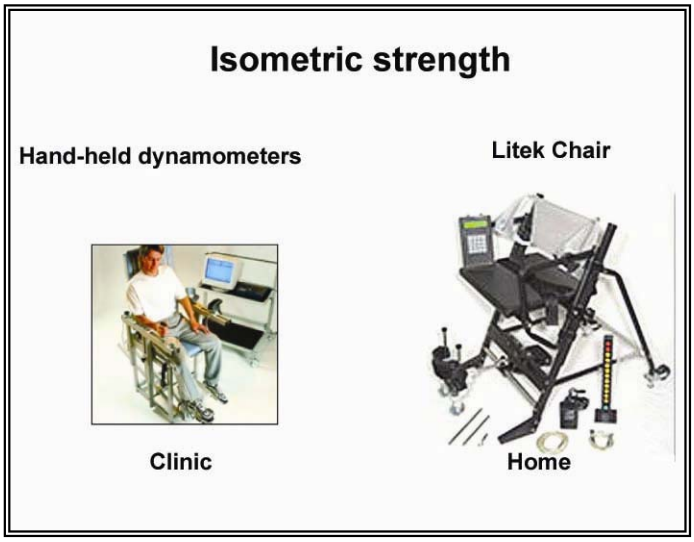
When the measurements start to require active cooperation, then you start to get missing data, functional performance, strength testing. And the missing data is actually informative, because it tends to be people who have difficulty with the measurement.

The last group of measurements, I think, can be designed so that you have a variable amount of missing data. And they test reserve capacity. And these kinds of measurements, I think, we are just beginning to understand how to put these into studies.

"...our experience, actually, was very population-specific. In Memphis, Tennessee, this was extremely welcomed. And in Pittsburgh, it was absolutely not accepted."

So my prejudice is that whatever it is that you're interested in measuring that fits with your hypothesis can be essentially affected whether you're in the home or you're in the clinic. And the things that I have stars next to are things where they require large instruments at this point, and people may have to come into the clinic. But there may be some surrogates available for at least some of these measurements. The x-rays of the hands, an x-ray box, can be taken into the home, and the retinal photograph, it's entirely that there's a portable machine to do that also.

So I think that with enough planning and with enough piloting to increase the precision of the measurement and to decrease variability and increase reproducibility, I think that you can move most of these things. And I've heard people talk about that. I've heard as you've discussed what's in your studies already. So this is just one example.



The way that you can measure strength really involves very small instruments that can be hand-held dynamometers. They can be instruments that are held by an examiner so that people have to push against them. But in order to remove the examiner issue, most people in a study of muscle strength will move to something that is a much more high-level isokinetic dynamometer.

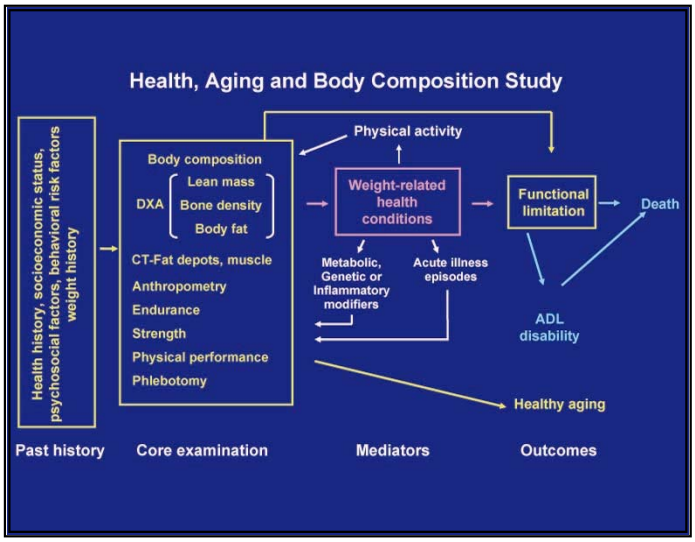
But there are at least two dynamometers that are available which measure isometric strength. So there's the good strength chair, which is a product from Finland, which is very good at measuring isometric strength both in the hand and the leg. And there's at least one portable chair in the health [AVC] study. We've actually used this and brought it into the home. It weighs, I think, about ten pounds. And it folds down very small and can be brought into the home and be used to measure isometric strength in the quadriceps, in particular.

Now, we talked about biomarker measurements, and these are some of the suggested specimen types. I've heard, again, all of these discussed here and I'm not going to go through each one in terms of collection. I didn't realize there was going to be another speaker about the sweat patch. I think that this is a very interesting technology. Initially, it was used for collection of data on drug abuse, and then bone markers. And since then, there's been a pilot study conducted at NIH in which analytes were isolated from the sweat patch.

And I think that the idea that you can put something on people that integrates measurements over a fair period of time is a very interesting idea. And it's an integrated measurement as opposed to the salivary measurement, which tends to be more time-limited.

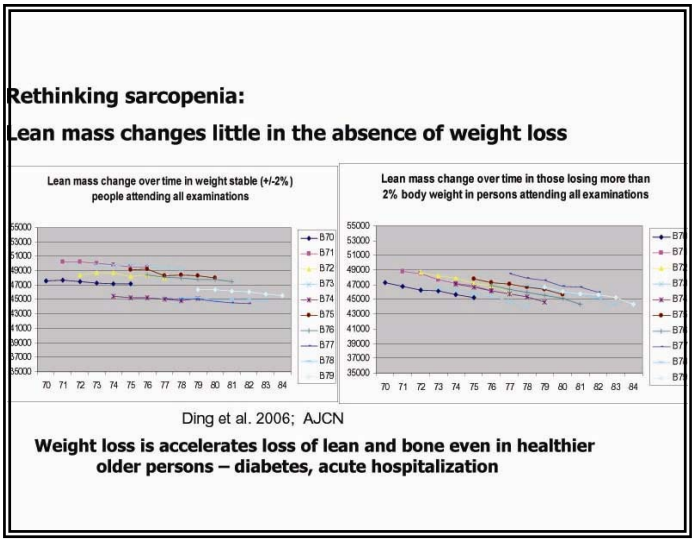
So for the rest of the time, I thought I would talk a little bit about my own work, which really is not measurements that can be brought into the home, but really tries to move beyond strictly looking at weight and height and waste circumstance as a way to look at issues related to weight and body composition in old age. And the reason that we needed this work was because there's a tremendous amount of controversy about whether being overweight is really bad for older people. Because as people get older, people tend to lose weight, and as people look across populations at older people, it looks like thinner people have a higher risk of dying, and heavier people actually are at decreased risk.

So we designed a study in which change in body composition was the common pathway by which weight-related health conditions contribute to risk of disease and disability. It's now been in the field about



ten years, and originally we started with 3,075 men and women. And everyone was at the same level of function; that is, they reported to us twice that they could walk a quarter of a mile without difficulty, and up ten steps without difficulty.

And this was the schematic that we used to organize the study, much like Dr. Lindau presented this morning, except that our core examination really focused on body composition measured in two or



three different ways: first with anthropometry; the second way is with dual energy x-ray absorptiometry, which is a very good way of dividing the body into lean mass, bone and body fat, but it doesn't tell you where the fat is and can't give you the actual muscle size; and then we use CT to tell us that information. We also measured strength and endurance and physical performance. And we linked this to weight-related health conditions and functional outcomes.

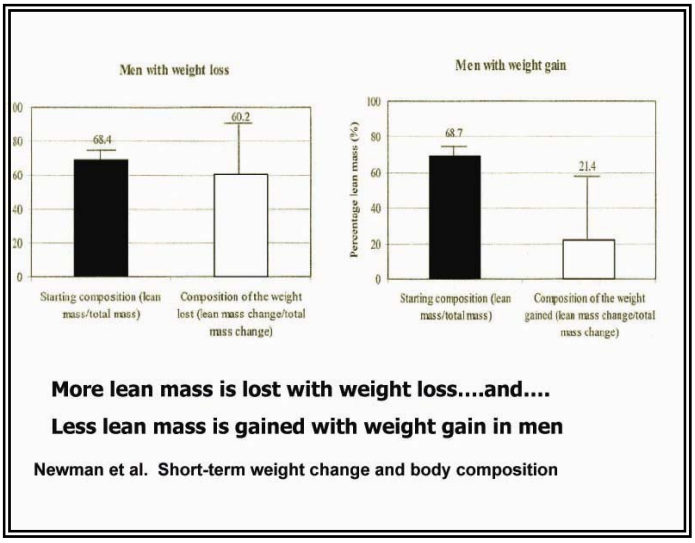
I do want to say that we tried using home examinations for function, strength, and collection of blood; questionnaire information; and information for accessing medication records. And I want to say that our experience, actually, was very population-specific. In Memphis, Tennessee, this was extremely welcomed. People had the examiners come to their home, and after that, they would come to the clinic. It was used very effectively as a response conversion tool.

And in Pittsburgh, it was absolutely not accepted. So people either talked to the examiners... If people wouldn't come to the clinic or couldn't come to the clinic, they spoke on the telephone but they would not take a home examination. And we're really not sure why that was, but I do think it is something to think about in terms of home examinations.

As part of the study, we tried to define a set of subclinical measures, and this is where the issue of biomarkers would come in. That is, we had the usual report of clinical history; then we did a series of physiological measurements to validate each one of the self-reports; and we carried this through and repeated it through the course of the study. And when we started to look at body composition -- let me just move ahead to this. So these are some of the things that we found using DEXA. And these are some of the reasons why you might be interested in getting something like total lean mass in your study.

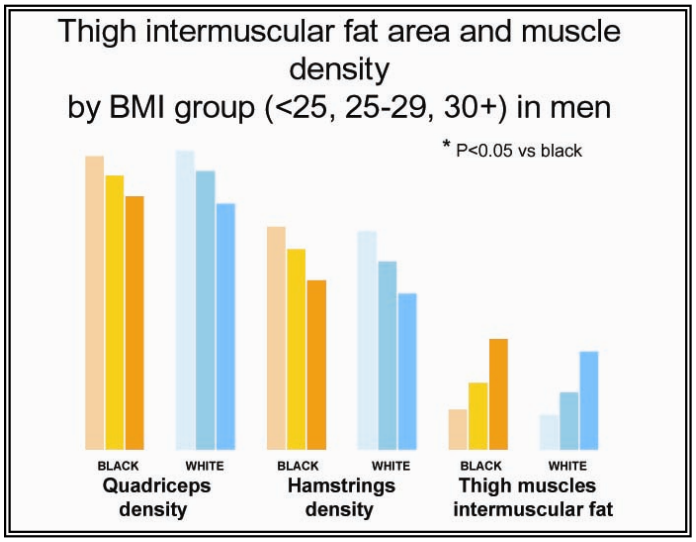
Now, you don't have to use DEXA. There are measures now from bioelectric multi-pole, bioelectric impedance measurements that will give you a fair estimate of the amount of muscle and fat that a person has, not only in the whole body, but you can also use segmental bio-impedance. But this is part of the findings that we have from the study.

So one thing we found is that it really doesn't matter; people who are of stable weight tend not to lose lean mass as they get older. A little bit of lean mass. But people who are losing weight are lose a lot of lean mass. And since about half of the cohort tended to lose weight as they got older, the main thing that was driving the loss of lean mass in old age was actually weight loss, not loss of lean mass.



We also found that the composition of the lean mass differed; that is, if you take men who lost weight, their start body composition for lean mass was about 60 kilograms, 63. And then it turned out that as they lost weight, they tended to lose lean mass, but if they gained weight, they tended to regain mostly fat. So they tended to not regain the lean that they had lost. We found also that the loss in leg lean mass did not account for the loss of strength. So these are the four

different groups in the population: white men, black men, white women, and black women. And in each population, the loss of strength was far greater than the loss of lean mass.



We also took our computerized tomography scans. And besides just accounting for the amount of muscle, we divided the leg into three different muscle groups -- the quadriceps, the ham strings and the adductors -- and we looked at this substance here, which is actually fat in the muscle. So those of you who haven't looked at computer tomography scans, this is actually...it looks like a steak. And yes, this is like fat. And then the actual muscle itself, in some people, is much more marbled than it is in other people. So you see this kind of little flecking which darkens the tissue. And that darkness is associated with increased lipid in the muscle.

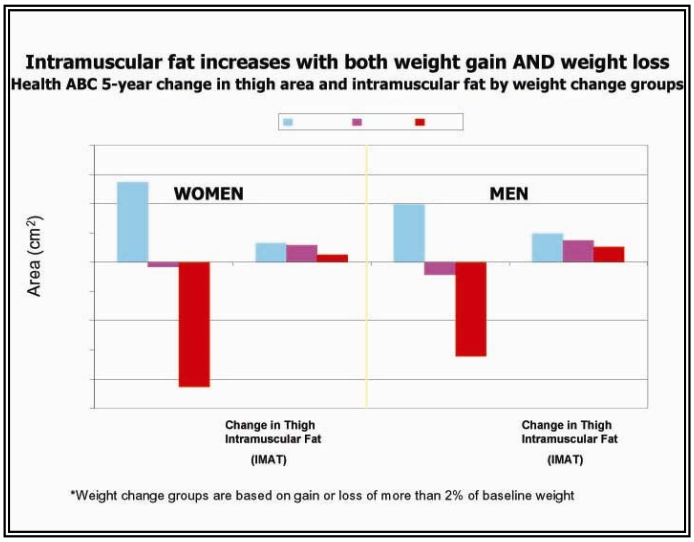
And it turns out that increased muscle in the lipid is a very important property. So one of the things we did is we looked at the relationship of that lipid in the muscle to BMI, and it turned out that regardless whether we looked at African Americans or whites, those people who were the heaviest actually had the highest amount of fat in the muscle -- which, this is a little counterintuitive because it means that it's the lowest the number here. But the lower the number, the greater the amount of fatty infiltration. And that's true for blacks and for whites, and it's true in the quadriceps, and it's also true in the hamstrings.

And then we also found that people who were heavier also had an increase in the amount of intermuscular fat, which were those large pools of fat that we saw.

When we look longitudinally, it turned out that muscle area, which is the amount of muscle you have, is important in terms of predicting risk of disability, but it's actually less important than the amount of strength that you have, and it's less important than the amount of fat that you have in the muscle. And even more interesting, these are data which a five-year interval. It turns out that whether you gain weight, are weight-stable, or you lose weight, that you actually increase the amount of that marbling, the large pools of fat within the muscle.

So these are all things that we're learning from having included measures of body composition in these studies. And the hypothesis that we're evolving, which I think plays very well into what Dr. Luke is going to discuss, is that people evolve to be very physically active. And essentially, their calories were always pursuing food, so that essentially, there was relatively little subcutaneous fat and a lot of muscle. And now, we've gotten to the point where most people are relatively...a little bit overweight and relatively physically inactive, which leads to a much greater deposition of fat and an infiltration of fat into the muscle.

And this happens increasingly as people get older because people are increasingly less active and less active in the ways that actually allow muscle to be increased. And there may be background changes in



the hormones that ____ muscle. But this is a 75-year-old farmer in Ireland, actually, a sheep farmer, who's walking up and down hills. And you can see that his muscle looks very much like this muscle. And I think if we had a better feel, anecdotally, from data around the world as to what energy expenditure was and can

document the maintenance of muscle, the maintenance of physical activity, I think that we would have a much better idea of what we need in order to remain healthy and well as we get older. Thank you.

[*Applause.*]

Amy Luke, Ph.D (Loyola University - Chicago)

Well, I would also like to thank the organizers for inviting me to speak on one of my favorite topics, which is doubly-labeled water. What I'm going to talk about are the principles of the doubly-labeled water method and how to go about the simplified protocol of the doubly-labeled water method. Then I'm going to show you a little bit of data from a study that's just been finished, and you'll see the kind of data that can be produced from using doubly-labeled water in a cross-cultural comparative study on physical activity in women in Nigeria as compared to women here in Chicago.

When we take a look at populations around the world, we know that chronic diseases are on the increase. And two of the exposures that, as epidemiologists and health professionals, we're interested in measuring are dietary intake or energy intake, and physical activity or energy expenditure. Unfortunately for us, those are two of the most difficult exposures to measure, and they've always been very difficult to measure. We're hoping we're getting better at measuring them, but we're still not presented with many good options for doing so, certainly not at the population level. I'd like to think, as someone who works in energy expenditure and energy metabolism, that we have a few better options for measuring expenditure than we do intake, and that we've been a little bit less contentious about it than those in the dietary intake field, but that may just be my bias.

So in dealing with physical activity in epidemiologic studies to date, we really haven't had a lot of options. Diaries and observation have been used, but they're incredibly burdensome: high participant-burden, high investigator burden. Accelerometery and heart rate monitoring, Dr. Stone has mentioned the Acti-Cal, are getting better in their methodologies and their equipment, but they're still far from perfect. There are still a lot of bugs to be worked out. But they are getting used more and more in epidemiologic studies. Obviously, the primary tool that's been used for decades is the questionnaire.

Now, I just did a very, very brief search and pulled a few examples out of the literature on the studies that have used physical activity questionnaires to look at the association between physical activity and mortality. And I just want you to take a quick look at the variability in how physical activity was defined and how it was measured, in just these six studies that have been published within the last eight or ten years. Whether it's amount of walking, number of stairs, leisure time physical activity combined with occupation, either frequency and duration, including intensity, in 18 categories or 21 categories, maybe including seasonality -- well, since this study was done in Finland, that was a good idea -- and then the time frame over which it was sampled, how you calculated it, and then how you measured it.

So you have to wonder, in all these cases, what are we really measuring? Is it true energy expenditure? Is it true physical activity? And more importantly, can we compare between studies? So what I'm going to talk to you about today is a method called doubly-labeled water. It's certainly not the cure-all and it's certainly not the answer for everybody, but it might be an option for some people. The doubly-labeled water method was sort of stumbled upon in the 1950s by a researcher at the University of Minnesota, Nate Lifson, and it was originally developed for use in small mammals.

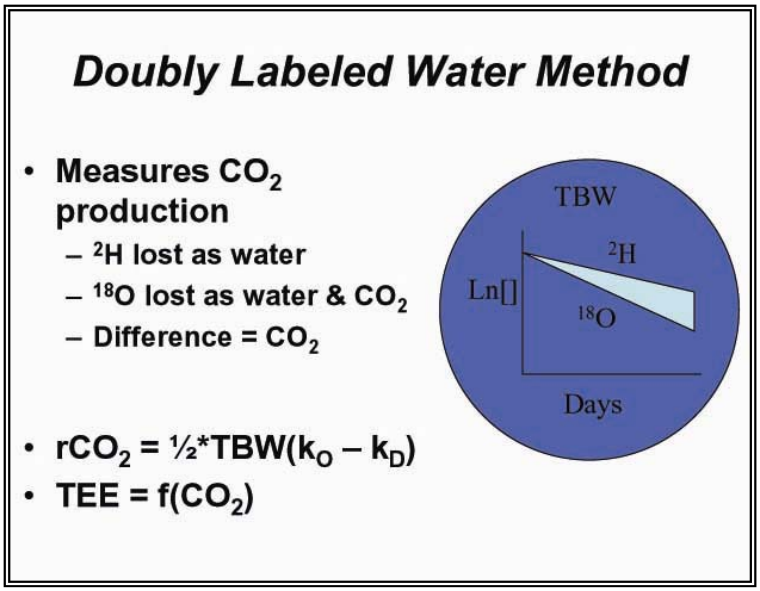
"...the strengths of the doubly-labeled water method are that it's objective. It's a noninvasive measure of free-living total energy expenditure. It's safe. It's been used by the WHO to estimate infant energy requirements. It has a relatively low participant burden except for that first day. It's relatively robust. It's accurate and precise, depending upon the experience of the research team."

Doubly-labeled water is an indirect measure of CO_2 production. Now as you may or may not know, CO_2 is the end product of energy metabolism in the body. It's the end product of cellular respiration. So when you go and have your basal metabolic rate measured in the CRC in your hospital, you'd be put in either a room or under a hood, and they would measure your oxygen consumption and your CO_2 production, and they would be able to back-calculate how much energy you produced during that period. So that's a measurement of your basal energy production via respiratory exchange.

The doubly-labeled water method provides an indirect

measurement of CO_2 production over your entire measurement period. And I'll just walk you through this. Basically, doubly-labeled water is deuterium and ^{18}O. In other words, stable isotopes of the two elements making up water, nonradioactive. The water is labeled with the two isotopes. You are given a loading dose of those isotopes. They equilibrate with your total body water, and then as you know, water leaves your body over time. Correct? The ^{18}O from the doubly labeled water equilibrates with the CO_2 in your lungs, the end-product of cellular respiration. Over time, the deuterium in that doubly labeled water that you've consumed leaves your body only as water. So it leaves it as urine, it leaves it as sweat, it leaves it as respiratory water.

The ^{18}O, or the oxygen isotope, leaves your body both as water and CO_2. So if you take a sample from the water compartment, your total body water, at baseline, and then again at the end of your measurement period, and determine the elimination rates of those isotopes, the differential in those elimination rates in the water compartment is an indirect estimate of the CO_2 produced over the whole measurement period. Basically what you have is another form of indirect calorimetry.



Typically, this measurement period is days or weeks. So you can give a dose of doubly-labeled water, keep the person in the clinic or under watch for four or five hours, because it takes that amount of time for the doubly labeled water to equilibrate with your total body water, and come back two weeks later and collect another urine sample, and you've got an integral measure of total energy expenditure over that two-week period.

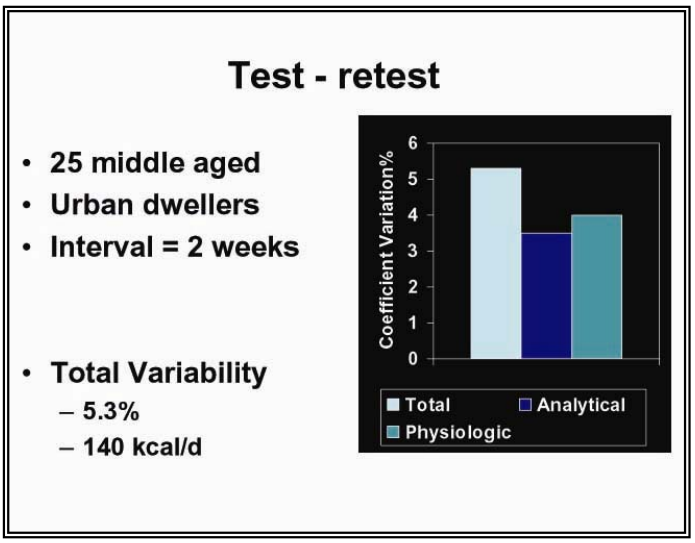
This is the original validation study conducted in Nate Lifson's lab, and it was done with mice. This is CO_2 production measured using doubly labeled water against CO_2 production measured with an indirect calorimeter. There was a very nice correlation. Doubly labeled water has subsequently been validated numerous times with respiratory gas exchange in humans. In the early '80s, my advisor at University of Chicago stumbled upon the same method. In 1955, it would have cost $20,000 to do the same experiment in humans, and that's because the cost of the isotope at that time was about $84.00 a gram. And the accuracy of the mass spectrometers at that time would have required 240 grams of isotope per person. So it was never even explored for humans.

By the time Dale [Schoeller] got around to it in the early '80s, mass spectrometers had improved to the point where it was more feasible. With today's instrumentation, it now requires only about seven to eight grams of isotope per person. Depending on body size, it may be as much as ten or eleven. And the isotope, depending upon supply, now costs between $20.00 and $40.00 per gram.

This is just to give you a sense of the reliability of the method. These coefficients of variation are not additive. But what I want to point out is the physiologic coefficient of variation. These are middle-aged urban dwellers, so that is why it's not a very high coefficient of variation; about three and a half percent. If you measure total energy expenditure younger people, you can get up to ten percent variability over time. And the analytic variability is what we'll come back to in a little bit. This was in Dale's lab, so it was a very good analytic variability. So if you have a very good lab, you can get very reproducible results.

The basic protocol is very easy. The method has a relatively low participant burden. You prepare a dose of the isotopes. After an overnight fast, you collect a baseline urine, because these are naturally occurring isotopes. So you have to measure every person's levels of the naturally occurring isotopes. You have the participant drink the doubly-labeled water. Usually, it's less than a hundred mL's. You then collect the next three urine voids over the next four hours as the doubly-labeled water equilibrates with your total body water.

For setting up your measurement protocol, there are a couple of ways to go about it. You can collect midpoint urine five to nine days after administration of the isotope dose and then collect an endpoint urine. Or you can collect two endpoint urines on the last day



and forego the midpoint sample. That entire sample time can be anywhere from nine days to eighteen days. As I said earlier, you calculate the difference between the elimination rates of the two isotopes, and that is a measure of CO_2 production and total energy expenditure. If you combine that with a measure of basal energy expenditure or basal metabolic rate, you can calculate physical activity by subtraction.

"Basically what you have is another form of indirect calorimetry."

Now I'm going to tell you a little bit about a study that has just been completed within the last year or so. The study uses participants from our international collaborative study of hypertension in blacks. It uses two of our populations, a small sub-sample: Igbo-Ora - Maywood Women's Study. It was a prospective study of the relationship between components of total daily energy expenditure and weight change in two cohorts of women who are at very divergent risks of obesity. We measured total energy expenditure using doubly-labeled water, resting energy expenditure by indirect calorimetry and calculated physical activity as the difference between the two. We also used a physical activity recall, a seven-day physical activity recall.

All you need to take a look at here is that the ages were similar, but as you can see, there were very different BMIs at baseline. But I want you to know that the Nigerian women were not malnourished as there was a mean of 30% body fat in the Nigerian women. These are not malnourished women by any means. This is total daily energy expenditure data. There are lower levels in Nigeria, but you'd expect that because they're smaller women, and you expect them to expend less energy. To compare between groups of different body sizes and weight, you have to normalize for body size.

This is activity energy expenditure in the two cohorts. When we normalize for body size, there's no difference in physical activity between the cohorts. This is our primary physical activity variable. This is another physical activity variable, using a different adjustment; again, no difference. This is a third physical activity variable; again, no difference at all. This was a completely unexpected outcome of our study. We fully expected the women in Nigeria to be much more physically active than our Maywood women.

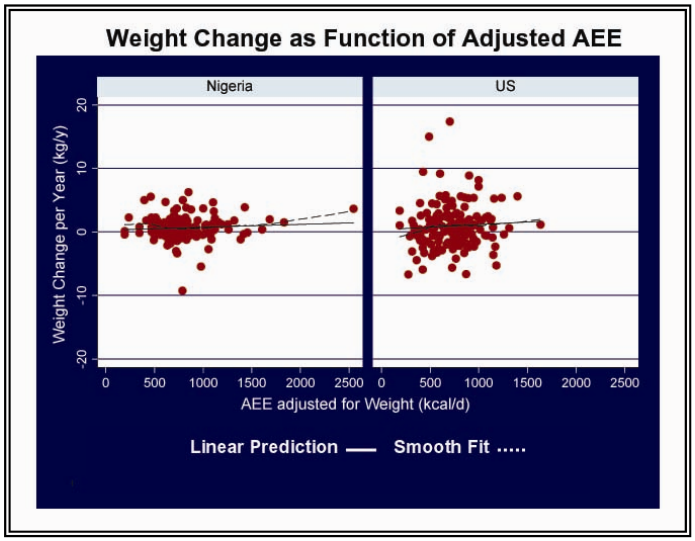
If we take a look at our weight change data, you can see that the women gained weight at exactly the same rates over the next three years. If we take a look at weight change as a function of physical activity adjusted for body size, there was no association in either site. Absolutely none. If we looked at it as a

Igbo-Ora – Maywood Women's Study
Activity Energy Expenditure – mean (SD)

	Nigeria (n = 153)	US (n = 179)
AEE (kcal/d)	755 (300)	800 (290)
PAL (TDEE/REE)	1.77 (0.26)	1.75 (0.21)
AEE adj. Wt (kcal/d)	803 (310)	748 (255)
Longitudinal Weight Data – mean (SD)		
Wt Change (kg)	2.3 (6.3)	1.3 (7.1)
Yr Follow-up (y)	3.4 (1.3)	2.4 (0.6)
Wt Change/yr (kg/y)	0.73 (1.9)	0.76 (3.7)

function of resting energy expenditure or basal metabolic rate, there was no association there, either. This table contains exactly the same data. We just broke it up by tertiles of weight change, and these are two measures of physical activity. So by tertiles of weight change in both Nigeria and the U.S., you can

see that the women who gained the most weight have the highest physical activity, but the association was not significant. These results are totally contrary, both to our hypotheses and our expectations.



And then as a caveat about questionnaire data, these are the associations between total energy expenditure using the physical activity recall and total energy expenditure by doubly-labeled water. In our Nigerian data, we had a positive association. In our Maywood data, we had a negative association. I'm not very happy about those results.

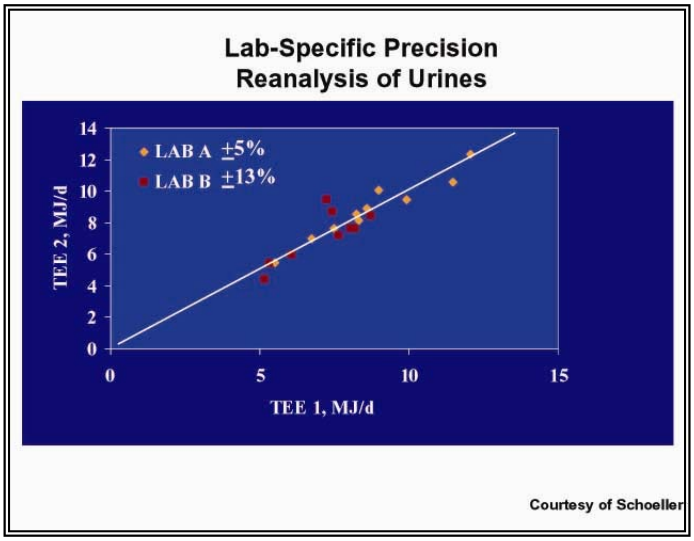
Just really quickly, this is the study that Dr. Harris talked about: the Health ABC Study. They used doubly-labeled water and indirect calorimetry to measure physical activity. They observed a very strong inverse association between activity energy expenditure and mortality. Through their questionnaires, they found the individuals with high activity energy expenditures climbed more stairs and worked for pay more. However, there was absolutely no difference in reported leisure time physical activity between those with high activity energy expenditure and those with low.

My point on this is that in sedentary societies, we have to be careful about what we're asking in questionnaires. I mean, if our questionnaires are focused on leisure time physical activity, I'm not sure that we're gathering the appropriate information to estimate physical activity expenditure.

To finish up, the strengths of the doubly-labeled water method are that it's objective. It's a noninvasive measure of free-living total energy expenditure. It's safe. It's been used by the WHO to estimate infant energy requirements. It has a relatively low participant burden except for that first day. It's relatively robust. It's accurate and precise, depending upon the experience of the research team.

"Doubly-labeled water is an indirect measure of CO_2 production. Now as you may or may not know, CO_2 is the end product of energy metabolism in the body. It's the end product of cellular respiration."

Limitations include the fact that it's expensive, but it's much less so than five years ago. When I first started the study, and the reason my N was much smaller with the doubly-labeled water sample than the sample I started with was that there was a worldwide shortage of ^{18}O and it cost $135.00 a gram. So the excess cost cut my sample size a bit. In addition, there's a delayed gratification regarding the data. It takes a while to run the samples in most labs. It doesn't provide information about physical activity patterns, but I'm not entirely convinced that questionnaires do, either. It can, however, be used in conjunction with accelerometers and questionnaires. It only provides energy expenditure data for the measurement period, but so does anything else that you use. You need well-trained field staff. That's not really difficult. We just finished a study in Cape Town, and the local PI



collected a urine sample in the rain. She got some water in there and diluted one sample, and we had to go backtrack and figure out what happened to that particular sample, but she learned.

And then you need an experienced laboratory for analysis, and that is more difficult. Just to show you, this is re-analysis of the very same urine sample, and as you can see, in Lab 1, you have a pretty good coefficient of variation, but in Lab 2, you have not so good replication of the exact same samples. So you have to be careful on where you have your samples run.

And other than that, I'd be happy to talk to anybody who wants to know more about this method of measuring total energy expenditure and physical activity.

[*Applause.*]

Short Presentations Part II: ***Influence of Culture on Health: From Bolivia to South Dakota***

Stacy Lindau, Introduction

It's my pleasure to introduce Thom McDade, who is a friend and a collaborator. He's been working with us on the National Social Life, Health and Aging Project, as I mentioned, and he works with many of you on your studies, as well.

Thom is Director of the Laboratory for Human Biology Research and Association Professor of Anthropology at Northwestern. He's also Associate Director of Cells to Society, which is at the Center on Social Disparities in Health, and a faculty fellow in the Institute for Policy Research. His work, and some of yours, is primarily concerned with the dynamic interrelationships among culture, biology and individual psychosocial environments across the life course.

And Thom is going to be presenting work today that is not from the United States. This relates to a recent paper that he published on ethnobotanical knowledge. It's my pleasure to introduce Thom McDade.

Thomas McDade, PhD (Northwestern)

Thanks to Stacy and Lis for the invitation to present again. In previous years, when I've presented to this group it's been mostly around methods implementation and development. I've always seen those as tools to answer interesting questions. So, it's a pleasure to talk to you about one of those interesting questions, and to show you some recent results from an ongoing study in Bolivia in which we're explicitly trying to understand physiology in relation to cultural and cultural processes in a population in transition.

A lot of us in epidemiology and anthropology are interested in culture, and we recognize the importance of culture as a determinant of variation in human health. And Stacy's very nice interactive biopsychosocial model had social, cultural context up on the sides of the hourglass there, which is pretty typical for a lot of social epidemiological models. But there's not a lot of elaboration about how to specifically model culture in relation to biology, and that's what I'm going to try to do today -- to talk about some new methods of doing some of those things.

We all know that culture matters to health, and human development. Medical anthropology has, as its central focus, the understanding of health and disease in the context of culture; it's the fastest growing sub-discipline within anthropology. But mostly what we know about culture and culture in relationship to health is ethnographic and qualitative in nature.

And that's because culture is, for the most part, a distributed, qualitative, shared phenomenon. The circles in this diagram represent individuals who are part of a group. And they are part of a culture that has a shared system of knowledge, ideals and behaviors that help inform their behavior and inform the interpretations of behaviors and actions and intentions of others.

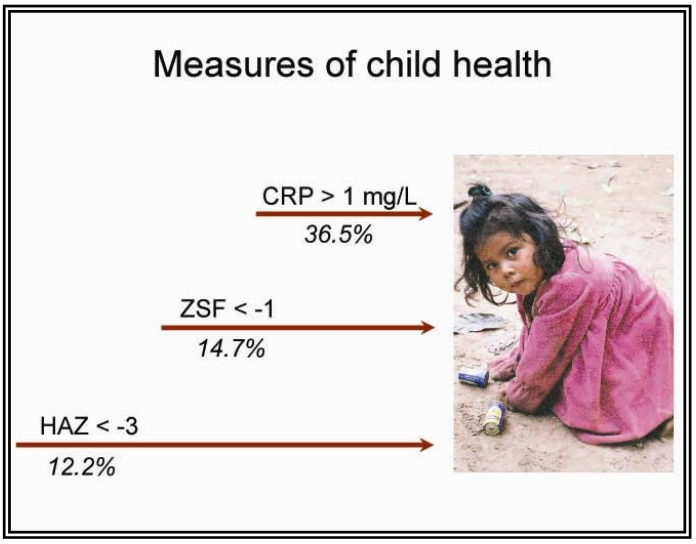
Culture is a difficult concept to get a handle on at the individual level, which is where we want to be when we're talking about health, for most cases. If we're looking at modeling a health outcome or a physiological outcome, at the end of the day, we need something that we can put into our regression or other quantitative statistical framework if we want to explain variation in an outcome in relation to culture as an independent variable. Culture is difficult to quantify at the individual level in ways that are not overly reductionistic or simplistic. But there are some new methods to do that, and that's what I'm going to talk about today.

So why the emphasis on culture? Culture is a key part of the human adaptive strategy; in fact it's partly what makes us human and distinguishes us from other, non-human primates. For better or for worse, humans are a fantastic evolutionary success story. For the past 20,000 years our population has exploded. We inhabit the widest geographic and ecological range of any vertebrate species on the planet. We live at 15,000 feet, we live at sea level, we live in the polar regions and at sea level in tropical rainforests, and culture is the reason that we can do this.

We rely on extra-somatic adaptations that allow us to harvest resources from our physical environment. We also rely on cultural processes to help us navigate our social environment. So culture is a key part of what it means to be human, and a key contributor to variation

in health, both across and within populations.

In our study in Bolivia with a remote Amazonian population, we see that local ethnobotanical knowledge is a key part of the adaptive strategy.



Ethnobotanical knowledge is simply what people know about the plants in their environment and what they're good for. This is something that historically has been very important for the human species, and, for many populations around the world, continues to be an important resource to manage a habitat, to eke out subsistence, and to use local plants and herbs for medicinal purposes. All these things have potential contributions to make to health.

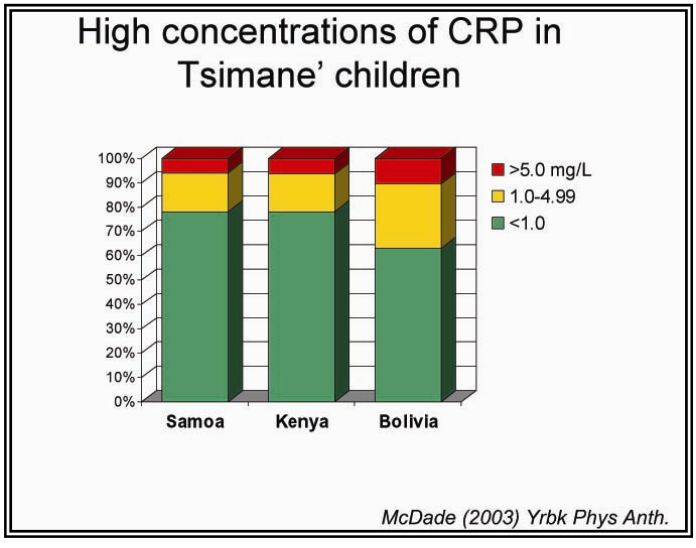
So it becomes a reasonable question to ask, does variation in ethnobotanical knowledge relate to individual variation in health? That's the central research question I'm going to talk about today. Does the ethnobotanical knowledge of parents relate to variation in the health of their children? And then, because we're looking at this in a population that is in the early stages of articulation with an emerging market economy, are these associations independent of other potentially confounding processes related to market integration and acculturation?

We conducted this research with the Tsimane', an Amazonian population of about 8,000 people in the Amazonian basin of lowland Bolivia. The Tsimane' inhabit a system of rivers, and there are no roads that access their communities. They live in small, semi-permanent villages of 50-200 people, and they are slash-and-burn horticulturalists, so they do small-scale agriculture and supplement their diets with hunting, gathering, fishing and a little bit of trade, but not much.

This is how we navigate to get to our field site, anywhere from four hours to four days, up and down the river to contact the various communities we're working with.

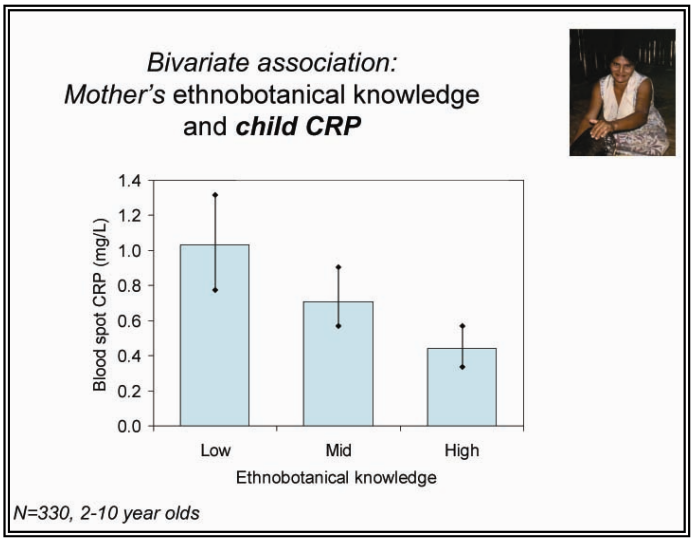
There are a lot of loggers and cattle ranchers that are moving in and usurping some of their traditional lands. And this is a problem, obviously, because it limits their opportunities for hunting and foraging and farming. But the Tsimane' also see this as, in part, opportunity. They can work for the loggers and cattle ranchers, and if you have a little money in your pocket, you can buy an antibiotic when your kid is sick. You can buy a rifle, which makes you a more efficient hunter, which improves the health of your children. So that dynamic, the push and pull of globalization, is really the central theme of our study.

In our study we are looking at three dimensions of child health. One is C-reactive protein. I'll have more to say about that in a moment. We also got anthropometric measures of height and weight and various skinfold measures. This is a measure of energetic status that changes on the order of weeks or months, and the amount of fat that you have in your body is an important source of fuel, particularly in this population, for fueling linear growth for children, as well as for fueling immune responses to infectious disease.



Height for age Z scores of less than negative three standard deviations indicate a long-term cumulative burden of under nutrition or a high burden of infectious disease. The results I'm going to show you are for two- to ten-year-olds, about 330 kids in the

study for these analyses. And about a third of them had elevated CRP. About twelve to fifteen percent were undernourished based on these categories.



We measured C-reactive protein in finger stick blood spot samples, using the same assay as the NSHAP study, as well as a number of other ongoing studies. But this is old school CRP--it's still the high sensitivity assay that everyone is using now, but we're interested in CRP as part of the acute phase response, as a clinically relevant indicator of pathogenicity, of the burden of infection that your body is trying to deal with.

Your body deals with a pathogenic challenge by up-regulating a number of pro-inflammatory mediators which initiate an acute phase response, leading to a number of symptoms associated with an infection, and also up-regulating production of C-reactive protein in the liver. C-reactive protein is a pattern recognition molecule which tags a lot of unfamiliar molecules in your body and helps clear them. It's a first line, quick response to an infectious agent that is very important in protecting you against infectious diseases while more specific, slower immune processes come online.

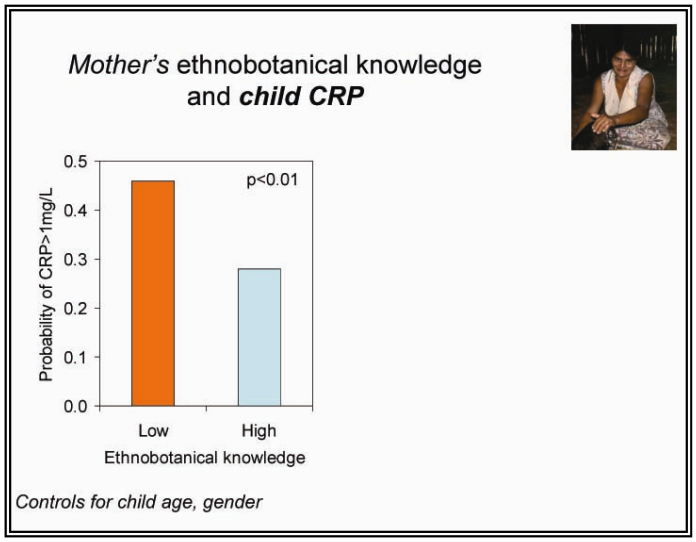
For our study we're measuring CRP as a biomarker of infection or immune activation, which does two things for us. One, it gives us an objective indicator of infection, whereas otherwise we would rely on observations of symptoms or reports from mothers about the symptoms of their children, and there are obvious recall biases and some observer biases associated with those reports. But more importantly, this gives us insight into sub-clinical activation of immune defenses, which are energetically very costly.

You or I may be dealing with an infection right now and not really be fully aware of it, and certainly might not report it if asked about it, but the physiology tells us something is going on there. And we particularly think that a high degree of upregulation of inflammatory processes is an energetic drain on this population that is leading to the high degree of growth faltering in the kids.

"A lot of us in epidemiology and anthropology are interested in culture, and we recognize the importance of culture as a determinant of variation in human health."

And just to show you a little bit of data on that, we did the same blood spot CRP method in three populations. Samoa, where I did my dissertation research, Kenya, and Bolivia. And you see that the proportion of kids who have some indication of upregulated CRP is much higher in Bolivia, in the rural situation we're working in, than it is in either Samoa or Kenya.

So then the question becomes, what is contributing to high CRP or other measures of impoverished health among Bolivian kids? And the one we're interested in right now is local ethnobotanical knowledge. The Tsimane' can tell us about over 400 plants in their environment. I probably could only name ten, I mean,



anywhere. And they can name over 500 uses for these 400 plants. So this is a key part of their cultural heritage. It is an important part of their local base of knowledge that they use every day.
This woman is weaving "jatata," which is used as

roofing material and as walls for their structures, so it's important for shelter. This woman is weaving cloth that is based on locally grown plants and is dyed with locally grown plants, and she's sifting through rice which is grown locally. These men have cut down a certain type of tree and are carving it out for use as a canoe to go up and down the river.

This woman is grinding up corn which is used to make chicha. What she'll do is she grinds up the corn, she puts it in her mouth, and then she spits it in this bowl right here, and then sets it aside for a month. And then the anthropologist comes knocking on the door and then that's what you get as your offering. It's supposed to be, I think, a welcome thing, but now that I think about it, maybe that's the way they try to keep us out. [*Laughter.*]

And this is a group of men who are doing what's called "barbasco" fishing, so they've diverted a part of the river, and this man has a bundle of twigs here which have been twisted and broken, and they're leaching out a chemical, a sap from that plant which changes the surface tension of the water, which, in effect, suffocates the fish. They float to the surface so even the kids can gather them up. It's a community way to use some plants to gather a lot of protein. So just a few examples of how plants are an important part of the daily lives of the Tsimane'.

How do we quantify this knowledge? The hypothesis is that people who have better knowledge of the plants in their environment and their uses will have healthier children. We're drawing on some methods from cognitive and cultural anthropology where we ask people to provide us lists of plants in their environment and what they're good for. We relied on experts, older people who have been around for a while, and they gave us these lists of plants and their uses.

And we took that list and we gave it to another group of people and asked them to match plants and lists. We subjected that to a modified form of factor analysis. And that gives us a sense of the degree to which this knowledge is coherent, the degree to which it's shared. And if the knowledge is shared, then we can call it cultural knowledge instead of idiosyncratic, individual knowledge.

It turns out this body of knowledge is widely shared, and we can quantify it and we can say this is the body of Tsimane' knowledge. And we take that list of plants and the list of its uses that the experts have told us about, and then we basically give mothers and fathers in the village a quiz. What plant is this and what is it good for? And then we compare their knowledge to the experts' knowledge, and it gives us a quantitative way to rank individuals with respect to one another.

"In our study we are looking at three dimensions of child health. One is C-reactive protein. We also got anthropometric measures of height and weight and various skinfold measures."

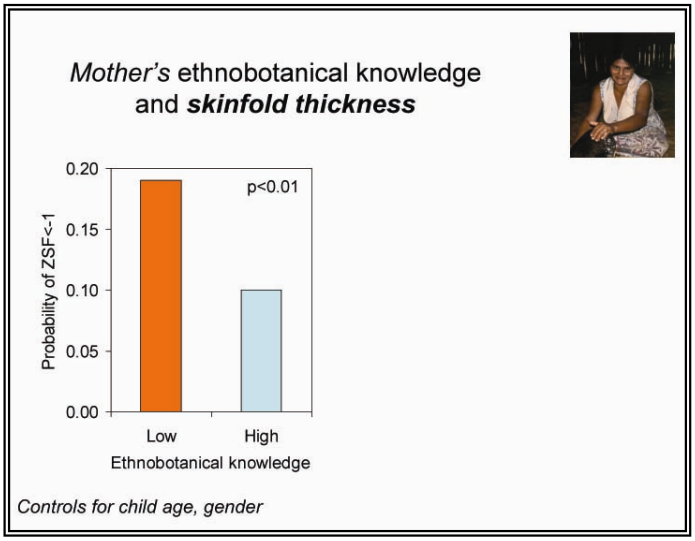
So we did that with knowledge of plant uses. We also did that with just basic ethnoscientific knowledge: what color is the Mahogany flower? When does it come into season? Just to get a sense of how tuned in people are into their natural environment. And then we also observed their actual use of some plants. We went to their homes, gathered plants, and looked at the diversity of plants that they were actually bringing in.

These are five different scales that have some very interesting theoretical differences among them. For my purposes, I lumped them all together and got pretty good coherence among these. And this is my approach to quantifying culture. So I have this group of people and I'm drawing on some theory that tells me what culture is, how to measure it, and why a mother's knowledge about her environment might matter to kids. I used some structured ethnographic methods, and I can locate specific individual women, with respect to her peers, and the degree to which she possesses this locally significant body of knowledge, and then I can use this to predict variation in the health of her children.

And we find some interesting associations. This is just the simple bivariate association, geometric mean, CRP concentration in children two to ten years of age based on the tertile distribution of the plant knowledge of their mothers. So kids with moms who have low ethnobotanical knowledge, their blood spot CRP concentration is about 1.0 milligram per liter, whereas it's less than half of that for mothers who have high levels of ethnobotanical knowledge.

Now there are all sorts of other things going on here, so in a logistic regression modeling framework, I

modeled the probability that a kid has CRP above 1 milligram per liter. And I chose 1 milligram per liter, which is lower than a lot of research on CRP for cardiovascular disease processes because we have



shown that having CRP above 1 milligram per liter is prospectively associated with reduced height. So it's a locally clinically significant cut point.

If you are a kid whose mother has low ethnobotanical knowledge -- this is one standard deviation below average -- then the probability that you have high CRP is about .44, and it's about half that if your mom has high levels of ethnobotanical knowledge. Now there's a skin fold thickness in the cut point here. It's less than one standard deviation based on American reference values. And again, if you have a mom with high ethnobotanical knowledge, the likelihood that you have low skin folds is about half.

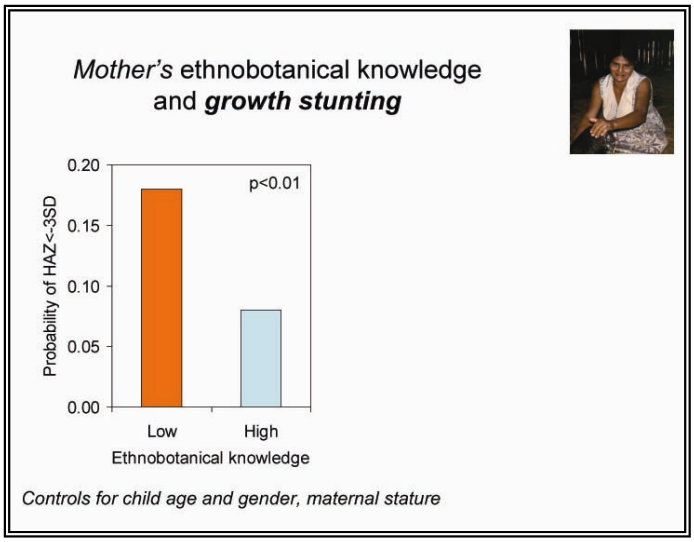
And similarly with growth stunting. About 12% of the kids overall were growth stunted. Eighteen percent of the kids were severely growth stunted if their moms didn't know much about the environment, and only about 8% of them were growth stunted if their moms had high levels of ethnobotanical knowledge.

Those associations controlled for the age of the kid and the gender of the kid, which were related to CRP. But what about these other processes related to market integration and health? We found that if a mom lives further away from the only town that's anywhere in that region, they had higher levels of ethnobotanical knowledge. Older mothers knew more about their environment, which makes good sense. Women who lived for longer periods of time in the same community had more ethnobotanical knowledge. Level of formal education, which was one or two or three years, was not related to ethnobotanical knowledge.

Interestingly, women who speak Spanish, which is an emerging language of importance to trade and commerce and through education, women who speak Spanish more fluently have lower ethnobotanical knowledge. And if they value their traditional Tsimane' culture, they have more ethnobotanical knowledge.

So these are the kinds of things we were worried about might confound these associations. This is the figure I showed you before. This controls for everything that came close, and basically the association is the same for child CRP. Same story for skin fold thickness. Same story for growth stunting. We found no evidence for confounding or mediation by processes related to market integration and acculturation.

So we found -- this is just a summary of what I just showed you -- that each standard deviation increase in a mother's ethnobotanical knowledge is associated with significant improvements in child health based on three different indicators. I was quite surprised by the strength and consistency of these findings. And based on that, I've convinced myself that we have something really causal here, subject to the limitations of non-experimental research.



So then the question becomes what are the mechanisms? And these are two questions we're starting to follow up here. One is that there may be some very specific pharmacological properties of some

plants in that environment that are effective in protecting the health of children. And then the other is that mothers who know more about their environment may just do better at feeding their kids, which also would protect them from infectious disease.

And then the other thing that I think is quite interesting and exciting about these findings is that we are documenting an adaptive function of culture. We are showing that culture matters to child health in a quantitative framework. And we worry about what's happening with the Tsimane' because we found that, for example, the closer you live to an emerging Western market town, speaking Spanish and not valuing your culture, these are things that erode ethnobotanical knowledge, and these are things that are only going to increase in their frequency in this population as they become more articulated with emerging market forces. So child health may take a hit in the near future for this population.

And the other thing that modeling culture in this way allows us to do is to compare the relative importance, in a quantitative, analytic framework, of an aspect of culture to other things we may be interested in, or other things that we know matter to health like educational status of mothers in particular, resources in the household. And it allows us to look at the relative contributions of these things and it allows us to look at how culture may actually act through some other things that we are interested in.

So these kinds of methods are…you know, when you go to Bolivia, culture sort of hits you in the face, but we have culture here, and we are doing studies in the U.S. where we are using these same methods to model aspects of social status, social relationships, sources of social supports and see how they are related to indicators of health in a birth outcomes study, and in a study in Maywood looking at adolescent stress. So I think these methods, which come out of anthropology, have a lot of relevance, and there's a lot of possibility for incorporating them in the United States as well.

[*Applause.*]

Q & A with Dr. McDade

Q: It strikes me that you could apply, in very interesting ways, the exact same method to the same topic of health literacy in this culture. And, I mean, the measures we have of health literacy, to my knowledge, are very bare bones kinds of cognitive measures, whereas understanding how to access healthcare and those sorts of things in our environments, it could shed a lot of light on health disparities. And I don't know if anybody's doing anything like that, or you're doing anything like that.

McDade: Not just literacy, but also what are understandings of the origins of health, or origins of disease? There is a postdoctoral fellow in preventive medicine at Northwestern who has used these same cognitive anthropology methods to interview different ethnic groups about understandings of the sources of cardiovascular disease. And my understanding of why I may be susceptible to a heart attack may be very different from another individual. And if you better understand why people think they might get sick, you should be better able to treat them or help them prevent disease. So yeah, very applicable here. Linda?

Waite: What you basically did was give these women a quiz.

McDade: Yeah.

Waite: Maybe what you're measuring is IQ, and maybe smarter mothers are just better able to, in lots and lots of ways, protect the health of their kids.

McDade: Yeah. This is the one thing that I don't have measured that I wish I had measured, and I don't know how to measure it. How do I measure IQ? That's the caveat to my statement that I feel like this is a real causal association, because that's the one thing that I feel like we didn't measure and couldn't measure. What I've thought about doing is giving these women a different kind of quiz and then following them up to see how well they can retain that knowledge, and to get at that sort of individual variation in intelligence question. Yeah, thank you. Yeah, Bob?

Bob: It would seem to me that…that same question had occurred to me, and it would seem to me that the thing to do might be to ask these women about their knowledge of some other aspect of the culture that has no obvious connection with children's health.

McDade: Mm-hmm, yes, yes. That's a good idea, and we haven't done that. We have done some literacy tests and we've done some tests of math skills, but

those are so tied to their experience with formal education that they're not meaningful measures of IQ, individual differences in intelligence.

Q: Just a subsequent question. It seems to me that as this is kind of cross-cultural that higher education is clearly correlated with improved health in our culture, and so maybe just using the marker of the botanical knowledge, in that culture that is the marker of either IQ or education in general, and although you can certainly see direct effects from the botanical connection, may just be a surrogate for education, as it would be in our society, a college degree or something.

McDade: So it's not in this case. So the formal education of these women is very limited, and it's not related to their level of ethnobotanical knowledge, but it could be in others. Globally, we know that the two most important predictors of child health are the education of their mothers and the mother's ability to manage household resources, and to have real, local economic development.

Those things really matter. They did not matter in this study. There was no variation explaining how it helped based on the formal education of the mother or the household resources that kids had access to or around them, which I think is interesting. And so that speaks to the importance, in this case, of these local cultural factors, which, this is a local, culturally relevant dimension of education that we would not pick up if we were WHO coming in doing a survey of Tsimane' health.

[*Applause.*]

Carl is going to take a second to load the take, so if there's one last question for Thom?

Q: The thing that was really intriguing to me was the fact that the people who were the most remote from the interface of modern culture were the ones who actually did the best, and that would really speak against the issues of intelligence and so on, because that would be not… I mean, the intrusion of the society is probably not a -- unless you're measuring intelligence strictly within the hierarchy of however the tribe lives, and I don't know what… I think that that's the other thing that you really didn't address: Is there a hierarchy in this tribe or in this clan, and could that affect the more remote living? It seems like you have to target the more remote living because that's the key feature that seems to be at odds with the cultural and educational influences.

McDade: Well, the more remote areas had better knowledge of local ethnobotany and use that more directly, but there's a lot of variation there. There were women who knew as much about plant use in the most articulated village than the most remote village. And there's not a lot of social stratification either, so that wasn't an issue. And I think, sort of following up on Linda's question that this may be the best local measure of intelligence, at least with respect to using resources to your advantage for the health of your children.

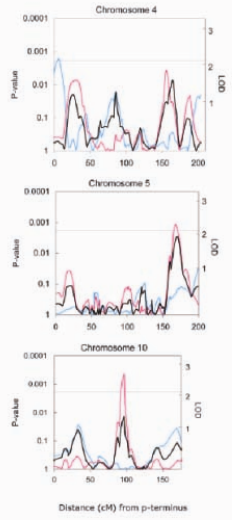
Lianne Kurina, PhD (University of Chicago)

Stacy Lindau

It's my pleasure to introduce Lianne Kurina, who is a Doctor of Biology, with her degree from Stanford. She's an Assistant Professor of Epidemiology in the Department of Health Studies at the University of Chicago. And Lianne is primarily interested in the relationship between distress and physical illness. She's currently doing work in the Hutterite population in the United States. And so a different population entirely from the one Thom is working in, but a unique and interesting population. And we're very pleased to have you here to talk.

Lianne Kurina, Ph.D.

I had always thought that my study population was pretty exotic, but it's difficult to compete with the groups that Thom finds. What I'd like to do today is describe our experience in setting up a study of stress and heart disease in the Hutterite population and hopefully give you a feel for some of the benefits and drawbacks of working with this unusual population, as well as the diverse types of data that we're collecting to characterize them.



Association Results

	D11S1981	D14S74
test	GTAM	ASHBD
location	11p (26 cM)	14q (78 cM)
allele	146 bp	301 bp
Direction of effect	(+)	(+)
Effect size	1.89 (female)	not estimable
Locus p (all)	9.2×10^{-5}	0.02
Locus p (female)	8.4×10^{-5}	9.1×10^{-5}
Locus p (male)	0.20	0.49

One of the central challenges facing those who do observational epidemiology is residual confounding. Even after we control for socioeconomic status and lifestyle factors, it's difficult to be certain that we're actually studying the effect of the exposure that we're interested in, and not a related confounder. So imagine, if you will, a study population where there is no smoking, where all the individuals share the same diet, the same day-to-day schedule, and socioeconomic status. The Hutterites are one such group.

The Hutterite population is a religious isolate that was derived from the Anabaptist movement in 1528 in the Tyrolean Alps. Religious persecution precipitated migrations around Europe until the 1870s, when 900 Hutterites moved to the United States and settled on three communal farms, which they call colonies, in an area that's now South Dakota. There are now more than 35,000 Hutterites living in the northern United States and western Canada. The demographers here will be very familiar with the Hutterites, as they are often used as an example of unrestricted fertility.

The Hutterites are a farming population, as I've told you. The top picture that you see is an aerial photograph of a Hutterite colony. These consist of somewhere around 80-100 individuals, maybe about ten families, each of which resides in a little apartment-like house within those larger continuous structures that you see. Hutterite culture is very gender stratified, so the men do the farming and deal with the livestock, and women cook, clean, take care of the children, and tend the vegetable beds.

And what you can see in the lower right picture is the children's dining room. So adults and children eat separately. The children eat together, and the adults eat together. Men and women sit at different tables, and actually the order of the seating is by age. And so these people, all their lives, with the exception of new people moving in, sit next to the same people and across from the same people every day. And the Hutterite dining rooms are extremely quiet. I mean, I think, you know, they're talked out, I guess.

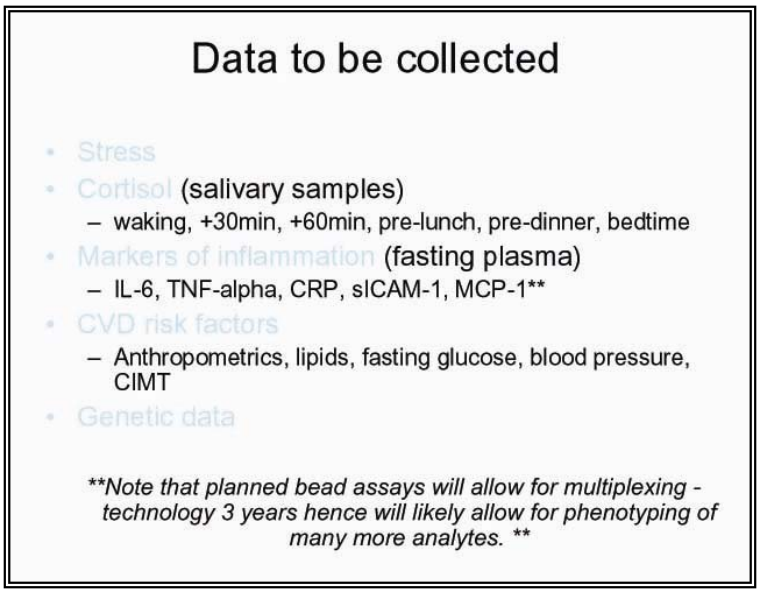
[*Laughter.*]

There's not much to say, really. The Hutterites differ from their Anabaptist cousins, the Amish, in that they use modern machinery, and these farms are really impressive operations. They often have foreign visitors coming in, for example, to tour their hog barn facility, to see how they achieve such high production and such low diseases rates. They have a really skilled work force that they've trained at a relatively low cost.

The one thing that I think I really want to convey is the isolation of the Hutterites. If you can imagine that you're on a small country road in rural South Dakota, exclusively lined by farms so that you rarely see houses, and then you turn off that road into a much smaller road and drive for a long time, you may come upon a Hutterite colony.

The first time that I visited there, after dinner we were walking to somebody else's house in the evening, and I saw this small child, a little boy who was maybe three, all by himself. It was dark. It was like 8:00 at night, and he was just trucking down the sidewalk, as little kids do, by himself. Coming from Chicago I was really struck by this, by what a safe and secure environment this is, and how different their lives really are.

So I was introduced to the Hutterites by Carole Ober, who's pictured here with a good Hutterite friend of hers. Carole is the person not wearing the head scarf. So Carole's been working with the Hutterites of South Dakota for over 20 years now, mainly studying the genetics of asthma and fertility. And because of her hard work, there's now a wealth of genetic data about the Hutterites, making this group of 700 individuals one of the best characterized genetic study populations in the world. And as many of you know, the large genome-wide association studies are really getting going. The Hutterite paper should be important in that literature.



A decade ago Carole also collected a bunch of phenotype data from approximately 700 individuals living on ten colonies, a subset of which I've listed here. And so Carole and I met because of the morning serum cortisol. I had worked with cortisol in the past we got talking, and she said, "Well, we're going to do a genome-wide screen." This was with a much smaller set of markers than they now have. So, we published a paper reporting the results of this screen in 2005. And what you see here on the left is the linkage results and on the right the association results.

One of the really striking things about our findings was the sex specificity of the results. This is joining a growing body of literature showing that it's often the case, even if you have a phenotype that's not particularly different for men and women, that the genetic influences on that phenotype could be sex-specific, and people are trying to work out why this might be the case. It's been observed in animal models for a long time, and now the evidence is growing that the same thing might be at work in humans.

The other thing that was kind of striking was the effect size, which means that, for this particular allele, the women -- only women -- who are homozygous for the 146 base pair allele had nearly double the level of morning serum cortisol compared to women who had no copies of that allele.

Now, it is a little hard to know what to do with morning serum cortisol. There are a lot of cortisol experts here. These fasting serum samples were collected within an hour of the individuals waking up, so it's saying something about the morning cortisol awakening response; we're just not exactly sure what.

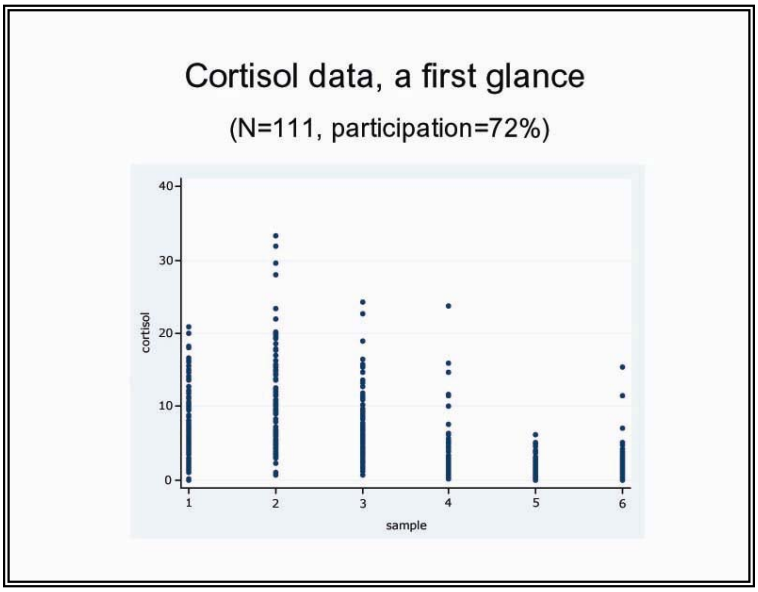
So when Carole said that she was thinking about going back into the field, starting this past winter, I started thinking that it would be great to have a more concrete cortisol phenotype. Then, we would have a chance to revisit – I guess really visit for the first time – a longstanding interest of mine in the relationship between stress and heart disease.

What one realizes is that with the Hutterites, you can have an opportunity to study the biological processes that might mediate the associations observed between stress and heart disease in the general population without worrying about confounding as much. So the environmental homogeneity is such that, if we observe relationships between non-acute stress and cortisol or inflammatory markers in the Hutterites, they'll be easier to detect if these relationships exist because of

reduced environmental noise and that they'll be better estimated because any relationship that we do see should be more substantial than a result of confounding.

Now, of course, the beautiful thing with this group is that we can also look at, and control for, in a sense, genetic influences on our potential biologic mediators of interest, and, if we have the power, to look at stress-gene interactions. And, finally, by collecting a new set of data on cardiovascular risk factors, we could look to see whether the morning serum cortisol collected ten years ago actually predicts change in cardiovascular disease risk over the past decade.

Okay, so these were the data that needed to be collected in order to put together some of the pieces of the puzzle described in the last slide. So how do we go about doing this, collecting what really is clinical data in rural South Dakota on these colonies? (We don't have the nifty NHANES vans.) Luckily, Carole is an expert at conducting field trips to South Dakota, but it was still an enormous amount of work, as anybody who collects data can attest to.



Before we could do anything else, we had to get permission to do these studies from the colony minister. This is the practical and spiritual leader of the colony. And further, especially in a study of stress and heart disease, we really had to get them on board with the notion of assessing stress in the Hutterite population. So it's not a big deal for the Hutterites to have asthma phenotypes done and to get their blood pulled.

But the stress interview that we constructed -- which you'll see more about in the next slide -- is not something that they encounter every day, and it was Carole's feeling that, actually, emotional discussions would probably be at a minimum among the Hutterites. She maintained that when a woman was pregnant, all the other women talked about it, but they didn't talk about it with that woman. So the kinds of conversations that Hutterites are having are different from the general population.

And actually, what we ended up doing was that, for each colony, any questions we thought might be sensitive and that the colony minister would be worried about, we queried beforehand to see whether it was okay with them if we asked these questions. In the second colony the minister said he didn't want us asking questions about alcohol consumption, for example, so we respected that.

Constructing protocols that could actually be done in the field in South Dakota, almost literally in the fields in South Dakota, was a challenge. And we were driving back from our initial visit talking about dry ice, "Where are we going to find dry ice?" You know, there's a little hospital in Mitchell, South Dakota, and we thought they might not be open, they might not even have that much dry ice, when Carole's research lab coordinator said, "Well, why don't we just call some bait shops?"

Maybe some of you know this -- I wouldn't think it's widely known in academic circles -- that bait shops carry dry ice so that hunters can ship their game back home. So it turns out that the Mitchell Wal-Mart has got dry ice, they're open 24 hours, and I got on quite friendly terms with Todd, the assistant manager to make sure that we had sufficient dry ice. So we sort of step by step managed these challenges, none of which, however, was as difficult as putting together a reasonable stress measure for the Hutterite population.

Let's just start with some highlights here. Clearly, an inventory of major life events used in other populations is not going to work. There's no divorce in the Hutterite population. You couldn't lose a job. No particular individual is under financial strain. And when an aging parent gets sick and there are caregiving responsibilities, these are big families. So the family size might be ten people or twelve people. I knew one

with seventeen kids. And so they get shared around. And I was talking to the Hutterite women kind of going through this interview and saying, "Did your parents get sick?" They said yes. I said, "Was it stressful to go give them care?" They said, "No, it was lovely. We visited with our sisters. It was a chance to see what recipes they were using."

So, the whole frame of aging and caregiving for aging people is extremely different, which will probably raise an obvious question in most people's mind, a question that was present in my mind at first: is this a reasonable population in which to do a study of stress? Are they stressed out at all? And I think the answer is that, in terms of traditionally perceived stress, as in things being out of control or uncertainty about what's going to happen the next day, this population was not stressed. But certainly, some of the other forms of psychological distress, we did actually observe reasonable variation, although I should say that the Hutterites do seem in better psychological shape, compared with normative data, than the general population. But I think we have enough range for traction.

So a few other things just to mention. There were questions and problems with understanding the perceived stress scale, which I'm sure many of you have worked with in your populations. When we asked the Hutterites the question, "In the last month, how often have you felt that things were going your way?" a number of them said, "You shouldn't always have things go your way." This is an un-American way of looking at the world, right, but clearly this question doesn't make sense.

And we ran into this time and time again where the notion of culture and the culture that they're inhabiting kind of butts up against what we traditionally think of as stressful or problematic in a person's life. I think they were mostly kind of perplexed/amused by the loneliness questions. It's very difficult to lack for companionship in a Hutterite colony, and they generally showed pretty high levels of emotional support.

So quickly, I just wanted to give you a little bit of a sense for what it was like in the field. We go to the Hutterite colonies in the wintertime because that's when they have time to participate in our studies. Farming chores are lighter, but of course it's very, very cold. But there's our fleet of minivans outside the butchering building. That was the space that we used for the clinic space in the second colony that we visited.

Here's what it looked like inside, and you can see a bunch of Hutterites. They moved from station to station around this very large space. You can see the men in their black work pants and suspenders. And here's some Hutterite children showing off their Ziploc Baggies that they carry around with them with little checklists so that researchers can say, "Oh, you need to go to this place next, you need to go to this place next." Okay, and this is just a data teaser because we're still starting analyses.

"And we ran into this time and time again where the notion of culture and the culture that they're inhabiting kind of butts up against what we traditionally think of as stressful or problematic in a person's life."

Yeah, it makes it look even more different than it is. But it's a plot that I know that many of you are familiar with. We've got cortisol levels in nanomoles per liter on the Y-axis; on the X-axis, the samples. We had them collect salivary samples at waking, 30 minutes after waking, an hour after waking, before lunch, before dinner, and bedtime.

It looks pretty much the way you expect, with one really big difference, which is that these cortisol values are low. So in comparison with the values, for example, presented by David [Almeida] last year from the MIDAS study, the mean levels at waking and thirty minutes following waking are, in the Hutterite population, about half of those compared with this other adult population in the United States. And this was true when I was kind of glancing around the literature for other populations as well.

So this is pretty striking. It could be consonant with the fact that they seem to be experiencing less psychological distress. But one thing should be clear, which is that these low cortisol values are not protecting them against heart disease. So obesity and abdominal obesity are really common in the Hutterites. We were shocked by the fasting glucose values, with the number of people with diabetes and pre-diabetes. So anyway, we're looking forward to running more analyses, but focusing more on building the sample. So

in the next few years, we'll be visiting another eight colonies to try to complete the sample and test these questions in a really rigorous way. We'll end up with about 800 individuals at the end.

And I should say, before I close, what a pleasure it has been working with the Hutterites. They are extremely warm and enthusiastic study participants. It's been a lot of fun. So let me just acknowledge Carole Ober, who very graciously allowed me to set up the study within this cohort that she's developed over the years, my research team for working so hard, and funding from The American Heart Association. And that's all. Thanks.

[*Applause.*]

Q & A with Dr. Kurina

Q: Do you know how many Hutterites leave the colonies? And might you be seeing low values of stress because only those who are very self-selected work?

Kurina: It's an excellent question, right, like the people who are happiest being Hutterites might remain Hutterites. So the answer is yes, that Hutterites do leave. Not in large numbers. Most of the time they come back. It's extremely stressful. That was one of our key life events that we ended up asking about because that is one of the most stressful things that can happen to them. And I think it's a fair question.

I only know anecdotally about a few cases, and I'm thinking in particular, most of the time, it's men who leave, but recently a couple of girls have left. Girls, I mean, late teens, which has been really distressing for the Hutterites. And I think in those two instances, knowing those girls, that it was sort of frustrations with their lot, in a sense. Education ends at eighth grade, and women don't get to

do very much beyond what's specified by their role. So in those instances, could it be the case that more stressed out, more depressed, more anxious people are leaving? Certainly.

Q: With that low cortisol, do you have longevity data or morbidity data?

Kurina: Only a little bit. What Carole has reported, based on pretty limited data, is that the Hutterites' life span is shorter than that of their rural peers. That's all I can tell you about longevity. And if you could see them, you'd sort of appreciate why. I mean, these are people who like to eat, and they like to eat rich food. As I said, abdominal obesity is really common, and heart disease is as well. So it's curious.

Q: The design sounds very much like a classroom student kind of design, a nested design, and probably should be analyzed that way, given the homogeneity within each community. So I'm wondering, have you thought about creating an objective stress level for each of the communities as opposed to doing this on an individual basis?

Kurina: I think that's a neat idea. We're only doing ten, so I go back and forth. There's certainly variability among colonies, and Carole will say that different colonies have different personalities. And when a colony gets too big, it splits, she believes, sometimes, along personality lines. If there is kind of intra-colonial conflict, you could resolve that by splitting up those groups.

On the other hand, their lifestyles are really close to identical. But we could think about differences and, you know, some colonies are richer than others, so to say there's no differences in socioeconomic status is slightly misleading, although compared to the rest of the world, it's not. But there could be subtle differences that they would be acutely aware of that might make a difference. So, yeah.

Lindau: If we want a break, we should probably take one last question.

Q: Are there any inbreeding issues here in the Hutterite community? And the other question is, while there's no social stratification, there is social differentiation by sex? And are there indicators that suggest different stress levels between men and women?

Kurina: Two great questions. So the answer to the first is yes. So if you were to take any two individuals from these 700 individuals that have been studied, they're related to each other, on average, as one and a half cousins. So this is an inbred population which, of course, was the motivation -- it's an isolate -- initially

for doing gene mapping in this group. You don't see the load of genetic diseases in the Hutterites that people have in the Amish, but I would say not yet. So now it seems to be the case that more cystic fibrosis is popping up, more developmental disorders, and there are population genetic models that suggest it will take X number of generations until these conditions start outing in an inbred population, and they might be starting to face that now.

Your second question is an excellent one, and I'm embarrassed to say that I haven't even done the sex stratified analyses for descriptive statistics for the emotional variables yet, but we definitely should.

Lindau: Okay. Thank you, Lianne. So why don't we agree to break for ten minutes, and start again at three o'clock?

[*Applause.*]

Panel Discussion: ***Paradigms in Genetic Biomarker Analysis and Integration***

Moderator: **Robert Hauser,** PhD (University of Wisconsin - Madison)

Vilmundur Gudnason, MD, PhD (Icelandic Heart Association Research Institute)

I want to start by thanking the organizers for giving me an opportunity to present some of the stuff that we are doing at the Icelandic Heart Association. And what I thought I would do, because they wanted me to discuss the homogeneity of the Icelandic population and how we are using that, is that I will give you an introduction as to what we stand for, and how the studies are, and what we are doing to basically come up with a way to utilize biomeasures to predict disease, and then give you a little glimpse of what we are doing, or what can be done, or how the genetics are. Of course, twenty minutes is not a long time, but I'm quite happy to discuss things with anyone afterwards.

So, basically, a little background: The Icelandic Heart Association is a nonprofit organization, it's a nongovernment organization, and it was founded in 1964. We do two lines of work: we run a cardiovascular risk clinic, and then we do epidemiology studies. In 1967, the Icelandic Heart Association founded this research institute and started the Reykjavik study, where the participants were all men and women born between 1907 and 1934. It was a total of about 20,000 people who have participated up to six times. So here is a schematic of the study, which was divided into six cohorts: one cohort seen every five years; one cohort not intended to be seen, which was used as a control; and the others, seen once or twice over the period. So basically, it's five full phases plus a sixth phase carried out in 1991 to 1997, but that concerned only people who were above 70 years of age.

So in here, we, of course, have these repeated measurements, then, of risk factors, mainly for cardiovascular disease. But of course, as you know, those risk factors turned out to apply to most chronic diseases. We have verified endpoints through serum samples frozen and stored from all stages. We have collected DNA since 1991, and we have real-time mortality and hospitalization data.

In addition, we have a myocardial infarction registry since 1981, started as a part of these Multinational MONItoring of trends and determinants in CArdiovascular disease (**MONICA)** surveys, and registration in the total population of all MI's in the whole of Iceland, of individuals aged between 25 and 74. All records are examined, including hospital records, autopsy records. We have about 11,000 in the registry. And of course, in addition, for the Reykjavik study, all cases are above 74 years of age.

One of the things we have done with this is to come up with risk charts or risk calculators based on the data that we obtained from these people. And this is very similar to the Framingham cohort or chart, and it's basically men and women stratified by smoking status, blood pressure, and cholesterol. This is used widely in our clinics, and it's actually available on our website for people to put in the risk factor and see how they do compared to their peers.

The important thing here is that when we look at European charts, which this represents, the European score project is comprised of a lot of smaller populations all over Europe -- part of the MONICA surveys, actually -- and they have come up with this risk chart. And when we compare the Icelandic or the Reykjavik score to the European score, what we find is that we have the same coefficients in Iceland as in European score project.

This is an extremely important point for those of us who want to tell people that what we find in Iceland is transferable to other populations. And this is a very important finding. So basically, Icelanders are like all other Europeans, believe it or not.

So, the newest phase of the Reykjavik study is a continuation of that study, which is the AGES Reykjavik study that you heard about this morning. And here, we are studying those who were still alive in 2002 and of course, including all of the available data that we have obtained through the years -- both what we have examined and what we obtained from the health system.

So this is a collaborative effort between the Icelandic Heart Association and the National Institute on Aging, where Dr. Tammy Harris and Dr. Lenore Launer, both of whom are here in the audience, are representing the

NIA. We are also collaborating with the National Eye Institute, National Institute for Deafness and Communicative Disorders, and the National Heart, Lung and Blood Institute. So, it's a very extensive collaboration.

And it's a big study that actually is drawing on the old Reykjavik study, so basically, we randomly recruited survivors from all the stages where they have participated and been examined in great detail. The recruitment started in September, 2002, and finished in January of last year. We had 5,776 participants, and we are actually embarking on another phase in this study.

But it's like I said this morning -- it's a study encompassing three half-days, so people come for three half-days. And I'm not going to describe in detail what we do in all these visits, but I just want to show you what we do with it, too. There we have this extensive imaging, so we have an MRI of the brain, and we have a multi-slide CT, which I will show you some research from, because we take scans of everyone's heart for coronary calcium, and then, of course, bone, muscle and abdominal scans for fat. And then we have ultrasounds of the carotids and we did some echo studies.

So basically, I just want to show you some results of the coronary calcium. This here is a classical coronary artery with calcium. You go slice by slice through this, extract the information, and convert it into Agatston scores. And here is one way we use it. So basically, this is a picture of a man's calcium scoring. Because of the extensive data we have on myocardial infarction and actually, also, interventions, bypasses and shunts, etc., we can group them according to whether they had a prior coronary event or not.

And the interesting thing here is that those who had a prior coronary event had much, much higher coronary calcium. The bulk of the people, of course, haven't had an event yet, although they were quite old. And we can see that for this man, coronary calcium rises with age, although here, at his oldest, the slope bends down a bit. But the main thing is that when we apply receiver operator characteristic analysis on this, then we come up with a differentional cutoff. That is this line. So there is a pretty good discrimination here. This is even more distinct with women. Those who have had coronary events have very, very high coronary calcium, and actually, the coronary calcium increases very steeply with age. And again, we can get a pretty good discrimination.

So basically what we have done is that we have used the data to come up with a risk calculator for the elderly based on the sixth phase of the Reykjavik study, and this is the first, actually, risk calculator for the development of cardiovascular disease in old people. So when we examine the first 2,300 that we recruited from 2002 to 2004, 44% of the 5% percent who had died after five years died from cardiovascular disease.

We look at, for example, CRP, and we can see that for the prediction of cardiovascular death, we see the blue line is according to our risk calculator; the red line is the CRP and it's not adding, I think, anything to the predictability. That's in line with what we had concluded from the Reykjavik study -- that the CRP is an independent risk factor, and it's really not adding to predictability. However, when we add coronary calcium, we can see that it adds considerably to the predictability of cardiovascular disease. So basically, we have now here a tool that we can use to add any risk factor, genetic or otherwise, to try to predict mortality from those risk factors.

So a little bit to the homogeneity of the Icelandic population. It is clear that there is a genetic homogeneity. For us, it means, really, that there's less locus and allelic heterogeneity. People have been focused on family-based research. There are genealogy databases where people have teased out a lot of families. And we published with a commercial company, a few years back, our findings on the five lipoxygenasa activating protein. And this has been replicated in some studies, but not in all.

And what we did, then, was to look at it in our own cohorts, and the 2,300 first participants. And we genotyped them for this FLAP gene SNPs, based on the MESA panel of cardiovascular genetics, or genetic SNPs, with some additional SNPs that we came up with for completion. So we found the risk increasing between two- and threefold. And the interesting thing here is that the haplotypes that we found associated with risk are not the haplotypes that were in the original paper. And that's, actually, quite an interesting finding. We then looked at the coronary calcium levels as a continuous trait, and this was not related to the original haplotypes either. And actually, the coronary calcium levels were inversely related with coronary calcium, and coronary calcium was not associated with the same haplotypes as for MI.

And when we look at this according to quartiles, we can see that for those in the top quartiles of coronary calcium it has a protective effect. There are fewer there with this haplotype. But the main thing here is that it's extremely rare; it's about a half of a percent. And basically, the conclusion of what I'm going to say is that we identified very rare haplotypes. And this, to me, means that family-based approaches are not necessarily the most successful approaches to identify common genetic variants, so then how useful is it for genetic studies?

Well, this here is a recent report in *Science* from Canadian, Danish and U.S. cohorts identifying this SNP on chromosome 9 associated with MI. It was replicated in the Icelandic cohort. It's actually interesting to tell you that it's based on the Icelandic Heart Association cohort. So basically, the important message here is that for this genetics, Icelanders are practically the same as other populations.

So I'll end here, believe it or not.

[*Applause.*]

Q&A with Dr. Gudnason

Q: I have two questions. I imagine you said this and I missed it. But were the haplotypes that you identified ones that you hypothesized *a priori*? Were those haplotypes you were looking for or were those ones you found by scanning the whole...?

Gudnason: These are basically identified by scanning. So these are the ones that came up. Those comparable with the haplotype A, which was the one identified in these families - nothing!

Q: Okay. And did you tell individuals who had these haplotypes that you found them, and that it appears they're associated?

Gudnason: To tell them what?

Q: Well, I guess that answers the question.

Gudnason: Yeah. We don't know what it means, so why go tell them something we don't know?

Q: Well, I think that's a question we all struggle with. So we don't know what it means; on the other hand, we publish papers with headlines that suggest that they significantly increase cardiovascular risk, so participants read headlines and might wonder whether they tested positive for that particular gene.

Gudnason: Yeah, you are absolutely right. But it's absolutely clear that none of these findings fulfill the criteria that have for appropriate screening for diseases, which actually, I have another slide for, but I'm taking that off.

Q: What's the relationship between coronary calcium and other measures of factors that might lead to coronary difficulties?

Gudnason: It's very clear that in our cohort here, and I presented that a few years back as a preliminary. We saw that clearly, because we had mid-life factors for these people. So it's absolutely clear that you can replace myocardial infarction with coronary calcium to find the same association as we've seen with these conventional risk factors.

Joseph Lee, PhD (Columbia University)

Hi. Today I'm going to talk about the way we approach gene mapping in Alzheimer's disease. And after going through the paradigm, I'll present two examples of the genes we recently discovered and published.

So before I forget, since I tend to forget, I want to give credit where it's due. The PI for the Alzheimer's Family Study is Dr. Richard Mayeux, and Ben Tycko is a molecular geneticist, and Yaacov Stern is a neuropsychologist. At New York State, Nicole Schupf followed Down's Syndrome patients for the past ten years or so. At U. of Toronto, Peter St. George-Hyslop is the molecular partner to this endeavor. At Boston U., Lindsay Farrer does quantitative genetics, and he has his own MIRAGE dataset and the Israeli-Arab samples. And in the Dominican Republic, we have Dr. Medrano, and in Puerto Rico, Dr. Jimenez recruits families on our project.

What are the goals of our research? The goal, first, is to localize; we want to figure out where the heck the gene is, whether it's in 14q, 19q or whatever. And then once we narrow it down to certain regions, we want to find out which one it is, because usually, region tends to include multiple genes, so we want to identify which one makes the list.

And once we identify it, we want to characterize how it influences the risk of disease. So, for example, in Alzheimer's disease, does it affect memory impairment? Does it increase the level of amyloid β? So we do that after we find the gene. And the goal is to better understand the biology of AD.

The problem is this: We know that Alzheimer's disease has a substantial genetic component, but to date, only four genes have been identified and confirmed. For these, everybody, even the competitors, believes that these actually do affect AD. And for the rest of the reported genes, your competitors do not believe those. So the problem is that to identify these genes that explain a small proportion of AD, you want to narrow it down from over 20,000 genes in the human genome to a manageable set of genes that you can study carefully.

So then if you look, there are about a thousand genes that are expressed exclusively in the brain. This is a large number of genes that you have to explore. So how are you going to narrow it down so that you can handle it given the limited resources -- financial or otherwise. And for late-onset diseases, particularly Alzheimer's disease, there are environmental factors to worry about.

It's not a single-gene disorder. So you also have to take into account education level, which seems to be protective against people's risk of developing disease, head trauma and so forth. And those things not only affect the disease itself, but they also interact with genetic factors. And then gene-gene interaction is an obvious possibility that we have to take into consideration. So there are multiple complexities involved here.

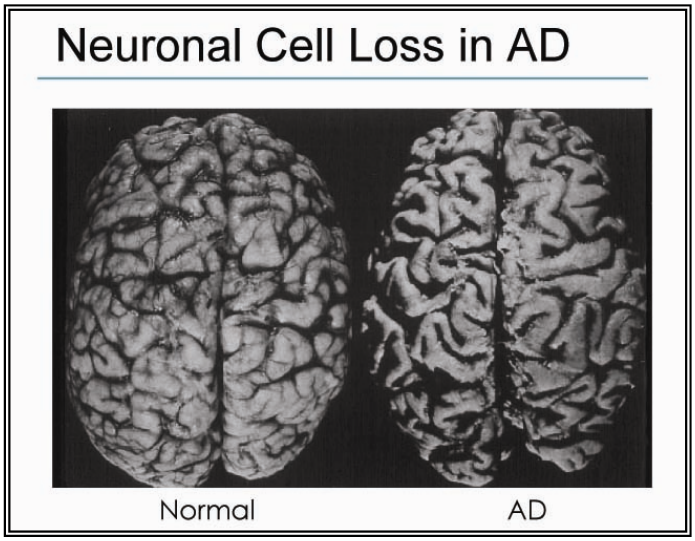
"The goal, first, is to localize; we want to figure out where the heck the gene is, whether it's in 14q, 19q or whatever. And then once we narrow it down to certain regions…"

So this is the outline of the way we approach this. Essentially, we start with a gene discovery phase of the study using the family study approach. We study large families with multiple affected individuals so that we can see if the disease is running in the family. And in some families from the Dominican Republic and Puerto Rico, there appear to be inbreeding where the relatives are marrying each other, so that these are particularly useful for recessive disorders. And I also do something called family-based migrant study, which I'm not going to discuss here today.

And then we collect a large number of the phenotypes that may help us understand the biology of the genetic contribution to Alzheimer's disease. And of course, once we identify a gene, we need to do functional genomics. Then we have a large cohort study from northern Manhattan that we've been following up for the last seventeen years with thousands of variables in elderly who are 65 and older. So from that cohort, we carry out a nested case-control study since we have limited funds.

To complement these studies, we also study Down Syndrome patients to understand Alzheimer disease. The reason Down Syndrome is interesting is because they have three copies of chromosome 21. This is

where the APP gene is located, and the APP gene is the one that produces amyloid β. So Down Syndrome persons essentially are swimming in amyloid β. Their brain may age much more rapidly than anybody else.



Between ages 40 and 45, in that five-year time period, you can see an aging process that is equivalent to about 20 years of aging in non-Down Syndrome individuals. So you can actually see -- even within my career, I can see somebody really getting old without waiting twenty years. So that's Nicole Schupf's study.

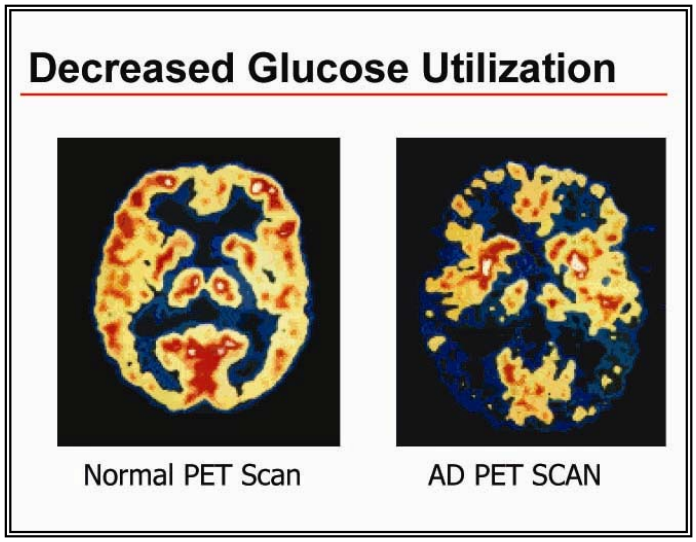
Okay, just to quickly go over Alzheimer disease. Individuals with AD show progressive decline of memory, language and orientation, motor function and so forth. Alzheimer's disease is characterized primarily by amyloid plaques and tau tangles, and that's why people talk about "plaques and tangles" when you talk about Alzheimer's disease. When these plaques and tangles start forming in the brain, it leads to brain cell loss and the brain mass starts to shrink, and then there's a change in blood flow and glucose utilization. So this one on the left side, you see a normal brain, and then on the right side, you see is an AD brain. And as you can see on the right side, glucose utilization is lower than the one on the left.

Risk factors known for Alzheimer's disease are these: age, Down Syndrome, having a positive family history puts you at a risk, and that means there probably are some genetic factors playing a role. African Americans and Hispanics tend to have a higher level of risk than Caucasians. In addition, environmental factors like education, hormonal status, head injury and so forth, modify risk of AD.

I am very much interested in finding new genes, which I call career-enhancing gene. To date, these are three deterministic genes that are transmitted in an autosomal dominant manner. On the other hand, APOE is a risk factor gene. When you have a copy of E4, it put you at an elevated risk of having Alzheimer disease. Lastly, there are lots of new genes that are popping up in the news. We found one ourselves. It's called SORL1, and it's on chromosome 11. Broadly, about ten percent of the cases are early-onset, whereas about 90% are late-onset AD. And for late-onset AD, only APOE has been confirmed.

I'm now going to talk about the Sortilin-related receptor (SORL1) gene. The study starts with the Hispanic familial AD. The study recruits most families in the Dominican Republic. This is the Dominican Republic and this is Puerto Rico. They tend to have large families, which makes it very nice for us to go gather an entire family at one sitting and study them. You cannot do in the United States, because most families only have one or two kids. Some are in a rural setting and some are in an urban area.

We looked at 228 families consisting of about 1,200 people, and about half of those are affected. And they have late-onset AD. We collaborate with Dr. St. George-Hyslop in Toronto, who has about half the number of families as we do, but his samples are of Caucasian background. In addition to these two main family study samples, we have access to case control samples of Caucasians, African Americans and Northern Europeans.



So this is the gene. It's a rather large gene. It has 48

exons, and it's about 117 kb. Because it's so big, we had to place a lot of **SNP**s to narrow down where the putative SNP may reside. Essentially, we genotyped 29 SNPs and then used a sliding window approach to walk across the gene.

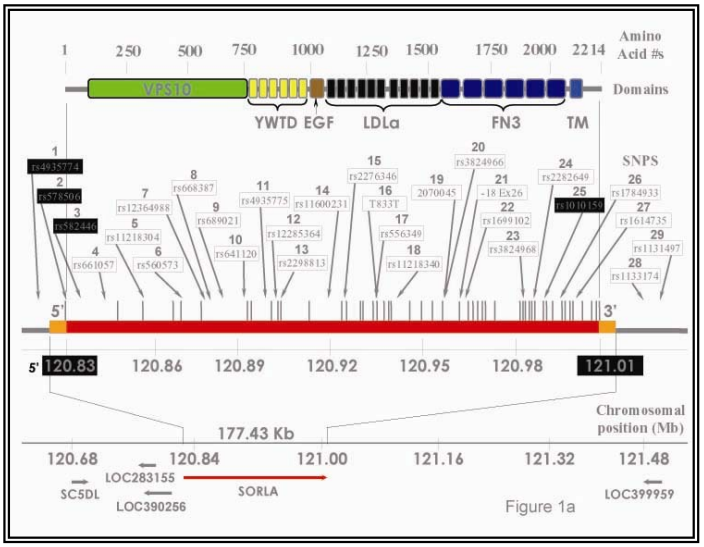


Figure 1a

One thing that I want to point out is that the linkage disequilibrium pattern in Caucasians is quite different from that in Hispanics. For example, in Hispanics, there are a large number of recombinations happening here, so the LD is really weak. So that makes it very difficult to study an allelic association in this area. You'll see that this differs from the Caucasian sample, and this pattern differs even more from the black samples we have. So it is imperative that one has to be aware of the different LD patterns across populations. This is going to affect the way you're going to be able to conduct your own study.

We performed a family-based study first. I'm not going to go into why it is important to do a family-based study, but we did a family-based association study because of its methodologic advantages. And in Caribbean Hispanics, we found SNPs C, G, C at SNP 8 through 10 to be associated with AD, and in Caucasians, we found SNPs T, G, T at SNPs 17, 19, 23 to be associated with AD.

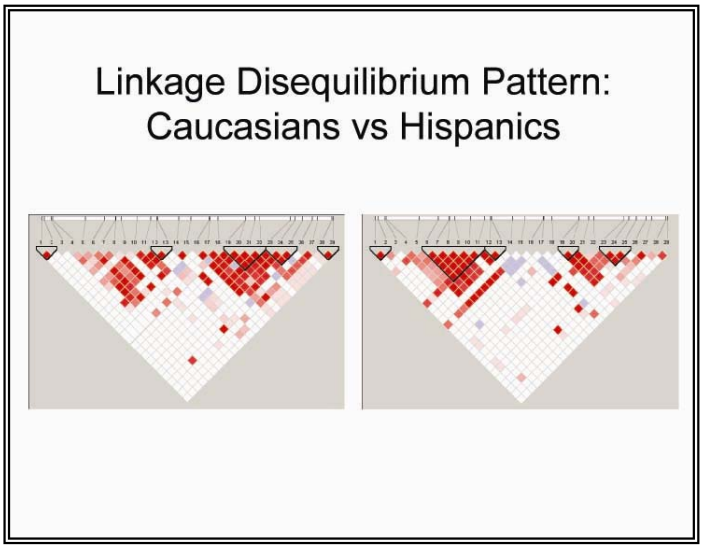
Then we replicated this finding in the Israeli-Arab samples, who are culturally isolated. This is Lindsay Farrer's samples. We also looked at Northern Europe case control samples. Interestingly enough, haplotype C-G-C holds up, and the G and T alleles at SNP 19 and 23 are holding up.

When we looked at the haplotype, C-G-C at SNPs 8 through 10 is significantly associated with AD in the Caribbean Hispanics, but nothing is happening in Caucasians.

On the other hand, if you look at haplotype C-T-T at SNPs 22-24 in Caucasian samples, the association is there, but not much is going on in the Caribbean Hispanics. You then slide the window one more SNP and look at haplotype T-T-C. This haplotype again is significant. Nothing in Caribbean Hispanics at this location. So it appears that two different patterns of haplotypes are segregating in two different populations.

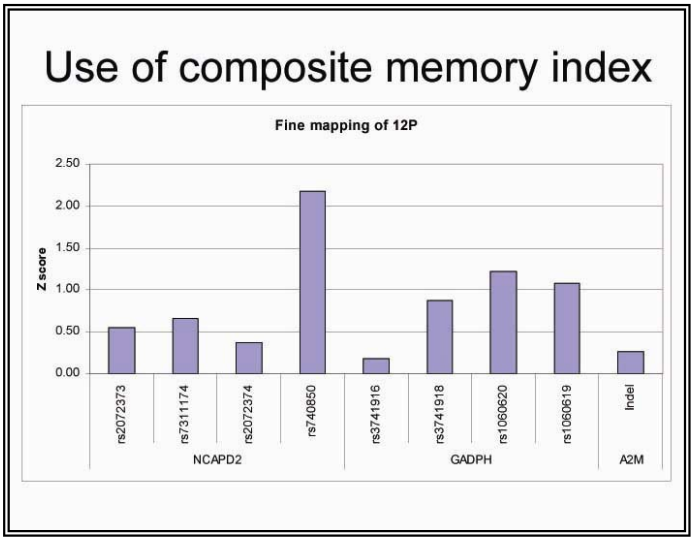
We looked further at the haplotypes in Israeli Arabs. They behave very similarly to the Caribbean Hispanics, so we were surprised. We have to do some population genetics and genealogical study, to see whether these Caribbeans have Middle Eastern contributions.

So we confirmed the TTC haplotype here at SNPs 23-25, and then the C-G-C haplotype at SNPs 8-10 in independent samples. But the reviewers still didn't like our findings. We got help from the Mayo Clinic. These have large Caucasian samples. They confirmed the T-C-C association at SNPs 23-25. So by the time we published this paper, we have genotyped over 6,000 people, and we have confirmation on two different haplotypes.



Following that publication in February, we published this paper in Northern Manhattan sample confirming the haplotype at SNP 23-25 in whites. Then we looked at the brain tissues and also confirmed the earlier findings. The brain tissues are important

because we want to be sure that the AD we see is actually pathologically confirmed AD. And then we looked at Down Syndrome subjects to see whether



having the haplotype reduces the level of SORL1, which in turn elevates the risk of AD by elevating amyloid β levels.
So things seem to be going in the right direction. I'm going to skip the functional genetics part.

Now, I'm going to present to you how we used memory as an endophenotype for gene mapping purposes. Chromosome 12p has been observed as a candidate region since 1997. Multiple groups have identified it, but there has not been one gene or one variant that explains that the linkage peak. So we decided to look at it. But why look at endophenotypes? There are many different pathways that can lead to Alzheimer disease. By looking at the phenotypes that are close to the actions of these genes, perhaps we can strengthen the relationship between the gene and the phenotypes we are looking at.

This is the bottom line. When we look at this gene, we see that Alzheimer disease is associated with a SNP for GAPDH, and 2 SNPs for CNAP1 here. These two genes were associated with AD. When we looked at memory scores based on a Selective Reminding Test, we see that there's some association with the GAPDH SNP and the CNAP1 SNP. To see whether this approach actually works, we looked at APOE as a proof of principle. Here, the E4 allele had a Z-score of 4.1 for AD.

When we look at the memory performance, you see that the Z-score for memory performance is -4.52, supporting the finding of an association we observed with AD. The negative score is correct, because, if you have the disease, the score will go down, so the sign should be flipped. You'll see that all the memory performance is associated with the E4 allele, and not with E2; E2 is protective. So this gives me the idea that whatever we're doing here at the chromosome 12p is working, because everybody believes APOE is a risk factor gene for AD.

Then, what we did was we took a bunch of memory scores and we came up with a composite score. We derived the composite score by looking at the prospective study from the Northern Manhattan Study. We had two peaks when we looked at the memory score before. When we used the composite score as the phenotype, this one was previously important and this one was also previously important. And in the composite score, this one comes out the winner. It is interesting to note that this gene is expressed in the brain; this one is not. So this finding even enhances the belief in what we're seeing here.

And then I conducted "unaffected only" analysis. Here, I do the same analysis, taking out the persons with AD. If the gene causes AD via memory, you should see the signal getting wiped away. And if the gene primarily affects memory, but less so for AD, then the association should remain the same or

Memory and candidate genes in 12p and APOE: Unaffected family members only

Gene	SNP	Total Recall	Delayed Recall	Delayed Recog	Benton Recog
	1	-1.03	-0.78	-1.40	1.18
TAPBP-R	2	-1.14	-1.42	-1.34	0.77
	3	1.00	0.54	0.86	0.44
	4	0.61	0.54	0.27	**-2.01**
	5	1.46	0.77	0.86	-1.71
CNAP1	6	0.16	-0.59	-0.60	0.52
	7	0.59	0.51	0.74	-1.08
	8	-0.43	0.11	-0.44	-1.79
	9	0.31	0.19	0.22	-0.45
GAPDH	10	-1.10	-0.92	-1.01	-1.21
	E2	0.78	**2.21**	1.17	-1.45
	E3	0.85	**2.40**	0.75	1.32
APOE	E4	-1.24	**-3.19**	-1.01	-0.86

increase. Here, you see that most of them get wiped away, suggesting that this gene is in the pathway toward the AD.

Now, if you look at the APOE, everything gets wiped out. But in delayed recall, this signal remains strong. This suggests that APOE may not only cause

Alzheimer disease, but it may cause memory impairment in an unaffected person as well. So it's a very interesting finding in that sense, at least for me.

In summary, we used multiple approaches to identify and characterize susceptibility genes. Multiplex families were used to identify the gene, and population-based samples were employed to confirm and characterize the gene. We used endophenotypes to enhance power to detect genotype-phenotype relations.

Thank you.

[*Applause.*]

Q & A with Dr. Lee

Q: That was a lovely presentation, and I think that you capture very well the huge amount of work that follows when you try to assess whether or not an SNP association in your data really is a true... And I wonder whether you think that this is going to be... Most of the people in the room don't necessarily have the level of data that you have now, but are just thinking about doing the kind of work that you've been describing.

And one question that I've been asking myself as I think forward to trying to do some of this is, a year from now, every genetics shop you go to, they talk about low-hanging fruit. Everybody is looking for the easy things to find. And what do you think is going to be the level of replication and information that will be necessary to publish in a year? What you've done right now in terms of replication and looking at the actual action of the gene and so on, I think that's at least the standard if not above the standard right now. But where do you think we'll be in six months, and where do you think we'll be in a year? Because it's relevant to what people have to do.

Lee: I cannot project what it's going to be like in a year, but with the number of papers reporting association, it is clear that efforts in this area are increasing. I'm sure it's going to be more competitive. For us to publish this paper, we had many difficulties, because we had such diverse populations. The reviewer wanted large Caucasian samples. Now, whether the biology only applies to Caucasians or it's only important when it's Caucasians, I don't know. But our study had over 6,000 people with a functional genomics, and we still had a hard time publishing it initially. So as long as you can pick the right reviewer, I think you're ok -- [*laughs*].

Q: [*First part of question inaudible*] ...in terms of setting up the cohort and doing the association studies, but from the time that you were trying to replicate, do you have any idea?

Lee: Using our samples, we're currently conducting a genome wide association. It costs about $500 per chip. And if you do any reasonable calculation, you probably need 500 cases and 500 controls minimally. That's a half million dollars; that's just to screen across the genome. Because we have done preliminary work with the family study, we know exactly what we are interested in, so we don't need to do that. But if you are starting fresh, and you have collected samples for the last fifteen years, and have a phenotype rich data set that you want to maximize, you probably want to do a genome wide association. That's going to cost you a half million plus.

Now, with Affymetrix and Illumina coming up with a million SNP chips... Of course, the chip cost is coming down a little bit. But the technology is getting fancier and fancier, and I don't see the cost coming down that much. But the thing is, once you've found that location with a p-value of $1x10^{-7}$, let's say, you need to do some more work to determine whether there's a variant that explains the phenotype. So that involves a substantial amount of effort to understand what the heck is going on there. So I don't know. I don't know how many people can afford to have a half-million dollar experiment with your ongoing RO1. I mean, it's just... Yes?

Q: I think my question is similar. I think I share some of the same thoughts. There are many of us in the room who are conducting studies that are collecting samples -- either blood spot or salivary seem to be the most common. Maybe whole-blood samples. The collection of them is complicated; it takes time in the home, and those are the costs. Their storage is an issue. We spend a lot of time -- we're going to talk more about the bioethical dilemmas we face.

Chances are, most of us are never going to use those specimens for genetic analysis because as you present, it takes a tremendous amount of expertise to make sense of data that might come from those specimens. So from your perspective -- and I think we all collected and NIH invests in us collecting this stuff because we

worry about the cost of not collecting it. But are we doing the right thing? And how do you think about which studies to go to when you want to do the validation pieces? And how do you get directed, is all this genetic data that we're collecting in this room of any use to you?

Lee: In my opinion, I think Dr. Gudnason from Iceland probably doesn't think too much of family studies. But different studies have different strengths. I believe family study does contribute important value. So since you have -- almost every one of you have a wonderful, rich population database, I think it would not be a bad idea to collaborate with the people who have family data.

And then as long as you talk about having uniform protocols and so forth, if the phenotypes overlap, then I think you have a much better chance of succeeding. So if the populations are reasonably comparable and the person who is pursuing families have done a good job of exploring the genome, and then you feel that there's something to be gained by collaborating, I think you have to collaborate.

For this one paper, we had collaborators from fourteen different institutions, and it was an international study. So I don't quite see how anybody can do this alone. And I think oftentimes, a lot of these genes have an effect size of 1.5 and 1.7. You're going to need a massive effort to put it together.

And I think if you have insightful biomeasures, it's going to even enrich your power to understand and dissect the genetic roles towards the disease of your interest. I think his study on Amazon would be cool because those Amazon people are essentially isolates in some sense, so that if you could start looking at those and then if you see -- and just like the Iceland -- more genetic homogeneity, maybe the Amazon people are an interesting homogeneous group of people.

Because I think that the simpler it gets, the easier it is to find the gene. After you find the gene, then you can look at all other things in a more complex way, I think. I like it simple. And if you can narrow down the complexity further down to simplicity, you may have something. And then you can ask all other smart people to figure out the relationships after.

Gudnason: I actually think you pointed out that I suggested family studies were useless or of little...

Lee: No, no, limited value.

Gudnason: The main thing is that there are rules of thumb. So for family-based studies where you have [first-degree] relatives, you practically have to have a genetic contribution of more than 50% to be able to pick it up. It's just a mathematical fact. People go for [sippers?]. You may be at 30%. The **deCODE approach** is basically second-cousin analysis. Go five, six meiosis, you pick up maybe a [tray-see-is] of 2, like in the [flap] studies. But for anything that is going to have a [high-set?] ratio or a ratio of 1.5 to 1.7, you need population-based association studies.

Lee: I did not say population association is not valuable. What I said was, you need to narrow down from twenty, thirty thousand genes to some reasonable things that we can manage. One of the best ways to cut down the number of genes that you have to look at is family study. Let me give two examples. One, the value of family study is that PS1, which explains less than five percent of Alzheimer disease, gave us the basis for understanding that plaques and tangles are important, and what the mechanism is by which that disease is progressing.

Therefore, by understanding the mechanism, we understand now how Alzheimer disease occurs, and then there are therapeutic means to address that. So it is perfectly fine to see whether it's a rare disease or not or it's effect in some small families. To me, that doesn't matter. Second thing is that if you have a lot of diseases where you see an odds ratio of 1.1 to 1.3, and because you've done a thousand cases and a thousand controls, well, what the heck does that mean? I mean, to me, it's more meaningful to say, well, this allele increases the risk twofold because it does something to it.

Because in the end, if you have a 0.002% elevated risk, I think that will be totally swamped by any environmental factors that you're going to see. So...

Gudnason: I totally agree with you. I think family-based studies are extremely important to identify pathways or possible mechanisms, but I think population-based studies are also going to be doing that.

Lee: Right.

Gudnason: We have a number of studies now in diabetes and cardiovascular disease which are basically genome wide association on populations, and they are coming up with new variants that will undoubtedly give us some information on the mechanism on the diseases.

Lee: Right, but I bet...

Gudnason: But both of them are important.

Lee: I'm not saying they're not important. As I said before, I use what -- I'm a pragmatist. I use whatever I've got and whatever I *can* use to my advantage. And to me, family study is intuitive. If I want to find a gene that does something to your disease, I want to see if that disease is running in your family. Because if it doesn't run in a family, it means that the gene isn't doing a heck of a lot, right? Because if a gene is running...if a gene causes a disease, it should run in the family. I mean, you should be able to see this. Of course it could be shared environment among family members. But if I see things segregating in the family, I feel a heck of a lot more comfortable than looking at anybody off the street. That's my...

Lainie Ross, MD, PhD (University of Chicago)

I will pass on your expression of gratitude for MacLean's support regarding this conference to Mark Siegler, the director of the MacLean Center for Clinical Medical Ethics at the University of Chicago. And thank you for inviting me. I'll represent the MacLean Center to the best of my abilities.

I don't need to tell anybody here what a biobank is, but what I want to look at, from an ethics perspective, is the changing paradigm in how we think about research. Traditionally, research with biological specimens involved collection by a single researcher or a group, and the samples were used for defined purposes to study a specific issue, and each research subject gave informed consent to use his or her sample for a specific project.

Biobanks offer a different paradigm, because the sample may be obtained by an individual entity who serves as a broker or intermediary for other researchers, someone who is not engaged in research. The purpose is often to develop repositories that can be used for many reasons and research protocols, often in numerous and diverse scientific areas, including future research activities not yet specified.

In fact, probably one of the biggest mistakes in biobank history was the way we decided to interpret genetic research using stored tissue samples. For example, consider the Framingham Heart Studies, one of the earliest databases, and the decision to say that permission wasn't given by subjects for genetic research. Of course it wasn't, because when the samples were collected, we did not know how to do genetic research. And so now researchers assume that they need separate consent for genetics and researchers attempt to make their consent forms very, very broad to ensure that future study methodologies will accommodate newer research methodologies and techniques.

And finally, biobanks seek to move beyond the one-study, one-informed-consent model to this notion of broad consent models so that the subject in a sense is really signing, in some ways, a blank check.

So what are the key elements of consent? They hold whether we're talking about one study or multiple studies and for however long we're going to store the samples. We have to ask, from an ethics perspective, is what are the risks, what are the benefits, what will the data be used for, will there be reporting back of data, which I got to hear a nice discussion about today, will the samples be given or sold to third parties, who will have access to the samples, to what extent are privacy and confidentiality ensured, and whose consent is necessary. And these are the topics I just want to go through today.

So what are the benefits of biobanks? Well, there's the potential to identify genes and modifier genes; the potential to increase our genetic knowledge and to study genotype/phenotype correlations; the potential to advance pharmacogenetics and the potential to promote personalized medicine. So many of these benefits are futuristic -- in a sense. They are not really going to benefit the individuals who gave the sample, but they might benefit future generations.

What are the risks? There's the disclosure of sensitive information. There's the risk that the subject sample will be used for research of which the sample donor does not approve. So in the old days, if you told me you want to study, for example, apoE and memory, I could say yes or no. But the fact is, you can also look at apoE to look at my cardio-vascular risk, to look at my risk of what happens after head trauma. It also turns out that apoE correlates with lupus and other autoimmune diseases. So it turns out that if we're going to use it for many different purposes, and I may approve, for example, of the apoE to be used for heart disease but not for memory issues and things of that sort. It could get more sensitive as we get into using samples collected for one clinical purpose to examine behavioral science research. There's also the risk that the sample will end up being used for clinical purposes about which the individual doesn't approve. So while I might be really happy that you've now found the association of memory and apoE, if individuals then decide to use it, for example, for prenatal diagnosis and termination of pregnancy, individuals may not be comfortable that that was the use to which their samples were used.

And finally, there is this notion of group harms. If we do find out that Icelandics are different than the rest of us or any other group, then it becomes a question of what type of stigma and discrimination will occur.

So the question, then, is what will the research be used for. The NCI Rand, when looking at the whole issue of best practices for biobanks, actually proposed what they called the 'tiered consent.' And you can imagine that people would agree to different statements, the first being the most modest: "I agree to my sample being used for this research project and then it'll be discarded." And that's not what we want when we collect these samples. They're expensive to collect, they're expensive to store -- we want to be able to use them for a lot.

So then the next level of tiered consent states, "I agree to my sample being used for any cardiovascular research but excluding genetic research," and that's how the old Framingham data were understood. Or we could say, "I agree that my sample be used for cardiovascular research, *including* genetic research," or I could just have literally a blank check and say, "I agree to my sample being used for any medical research," and then we might get into disagreements about what counts as medical research. For example, if you're looking for genes for intelligence, is that about medical research?

And then you get into nonmedical research and whether one can use the samples for forensic purposes. So since we're in an international audience, I just thought it would be important to tell a couple of anecdotes from around the world to understand the impact and the importance and the ethical issues that come about from biobank. So how many people know the story of the Anna Lindh affair in Sweden? Okay, so we have one hand up.

So this was an interesting. Anna Lindh was murdered and she fought off her assailant. She actually was running for Prime Minister, and the elections were supposed to be the next day. And she fought off her assailant, and they had DNA material underneath her fingernails. And they apprehended an individual who they thought had committed it, and he didn't match the DNA samples and had to be let go.

It turned out this was the second murder of a prime minister or -- the other one was a prime minister -- within two decades in Sweden. And the answer was, when Olof Palme was murdered in the late 1980s, they never found the assailant. They decided in Sweden that they couldn't allow this to occur. And they have, like we all have, newborn blood spots, which has every single DNA of every single individual ever born in Sweden. And they went to the Supreme Court in Sweden and basically said, "We want permission to go and match the DNA under Anna Lindt's fingernails to the blood spot."

So they located this individual, at least genetically, and now they just had to find the individual, who was eventually located and apprehended. This raises huge issues because those newborn blood spots were collected for public health purposes, maybe to be used in future for epidemiological purposes, but really collected to test for PKU, hypothyroidism, and maybe a couple other diseases. And because of this, the equivalent of what we would call the American Civil Liberties Union got involved and now there are new laws in Sweden which allows individuals to remove their blood spots.

Can you imagine who removes their blood spots? They're the privacy freaks. They're the people who have a reason to have it removed, in case their DNA ever got matched to...I don't know, a crime or anything of that sort. And this makes it, then, hard to do any epidemiological research that was possible a decade ago.

"We have to ask, from an ethics perspective, is what are the risks, what are the benefits, what will the data be used for, will there be reporting back of data, which I got to hear a nice discussion about today, will the samples be given or sold to third parties, who will have access to the samples, to what extent are privacy and confidentiality ensured, and whose consent is necessary."

And so that brings up the whole notion of how these blood spots are protected and ensured. We can talk about certificates of confidentiality here in the United States, which are certificates that are supposed to protect your samples from a subpoena. It turns out that Bush, in signing some of his documents allowing him to go to war with Iraq, has made a loophole so that he, too, could actually get the Supreme Court to open up any one of our DNA biobanks under the Patriot Act. And that's a really important issue for people who are worried about privacy and what you can actually promise your participants in your research studies.

Another really important issue is this whole notion of reporting back of results. I was actually thrilled to hear when the question was, "So, are you reporting back the results?" and the answer was, "What is there to report back?" It turns out that Children's Hospital of Boston announced last month they're going to have to provide access to preliminary results to the parents who enroll their children in the Children's Hospital of Boston biobank.

How preliminary? Well, that's going to be decided by a group that will include scientists, experts, and lay public. But their whole goal is to increase the number of people willing to participate, and also, they feel, to get rid of paternalism in research. And yet my fear, as an ethicist, is that this is going to give people information that has never been validated or replicated. And even if it were to then be found to be replicated and validated, we don't even know what it means.

And so what does it mean to tell a parent that your child has an eight percent increased risk over the general population within the next sixty years of having a stroke? And how will this influence what physicians tell them? I would tell the parent, "Well, the data doesn't really change anything." I would tell them to do what I would've told them to do even before I got that genetic information, which was to sleep right, drive below sixty, eat breakfast every morning and try to lead a stress-free life.

So how, in a sense, is this genetic information going to change anything? Well, it turns out that there are studies that are being done -- again, in Sweden but also in others parts of the Netherlands -- looking at HLA risk factors for diabetes. And what we have found is that some families will actually change their children's diets even before the development of diabetes, even though the researchers will tell them that will have no influence on your children developing Type 1 diabetes. So people *will* use the information in ways that we don't necessarily think makes medical sense. And so we really have to think about what it means to report back results.

Now, there could be some clinical benefit. If you learned about genetic susceptibilities and did change your lifestyle -- you started eating right, you started getting enough exercise -- but there also the clinical risks of misusing this information, or of doing things that may be harmful.

There's also the question of whether there'll be adequate genetic counseling available for all individuals. For example, this Harvard program has it so that you can pick it up off the computer. So there's no genetic counselor who's sitting there on your computer explaining to you what it means to say that you have an eight percent increase over the general population.

It does have potential benefit for the researchers because on the bottom of the Harvard screen, it could say, "By the way, we want to validate and replicate this study, so send in all your relatives." And so this way, it could have a real research benefit in getting clusters of populations when that will be beneficial for the research.

So the question, still remains, is it ethical to report back the preliminary results? And the Harvard researchers have basically said, "Research participants have a right to this information." And that may or may not be the case. The fact is, your consent form tells individuals what rights they have. Your consent form can say you're not going to get back results, and if I don't like that, then I don't have to enroll in your study. So, all of this really ought to be negotiated in the consent form. I would actually say that researchers have a responsibility not to report back individual results until we really know that the research can be validated and replicated, and that we actually know what it means.

So I would argue that what we want to do is wait until it achieves clinical usefulness. And at the University of Chicago, we've developed three different databases. One in the Department of Obstetrics and Gynecology, called CLIPP (University of Chicago Lying in Pregnancy Project), with Carol Ober, whose name was already mentioned today, that only wants to look at issues of infertility, miscarriage and pregnancy issues. There's one in medicine called TriDOM (translational research institute in the Department of Medicine), and there's one in pediatrics called KidsGene.

And at least in KidsGene, for which I was involved in writing or revising the consents, we're very clear that we will report back results if you want it only if we can say that it's going to be clinically useful, it's been corroborated from different samples, that there's an adequate understanding of genotype/phenotype correlation, and that there's some measure or

preventive something that can be done that will allow individuals to make sense of this information.

And, of course, you could also argue that researchers, before reporting back results, should ask for a second sample; this whole notion that mistakes are made with samples, and you would like to confirm that you're actually giving back the right data to the right individuals -- something like what we call here in the States CLIA (Clinical Laboratory Improvement Amendments) -approved labs.

So we have looked at the issues of what does it mean to consent and when will we report back results. The next issue is who will have access to the sample. And again, the answer will depend on who created the biobank and their purposes. Some of our main players are academic medical centers, which may be doing it for their own research, although NU-Gene here in Chicago -- that's Northwestern's large population database – has plans to sell samples to help fund the biobank repository in the first place. So if you call up and you say you want forty individuals with asthma, they could give them to you for a price.

Pharmaceutical companies are also creating biobanks. Advocacy groups -- this is a real interesting one -- Sharon Terry and the gene called PXE, pseudoxanthoma elasticum. And what she realized, and I'll be giving a couple more anecdotes, is that individuals gave samples and communities gave samples, and then researchers went and patented the genes. Then, all of a sudden, the tests that came out were now prohibitively expensive, and the community that gave the genes wasn't necessarily even able to afford to get the testing for themselves.

So Sharon Terry will give you a sample, but you have to sign a contract with her that she's going to be, for example, named on any patent. She was one of the authors on the article that came out when they located the gene for PXE. And she is named as one of the individuals that licenses the PXE gene, and so that the money can go back into PXE International and things of that sort. So advocacy groups are getting a lot smarter or savvier on how and when to use their power and the value of their collected samples.

And then there'll just be middleman, people who, for a fee, will collect and distribute samples for the researchers in a de-identified manner, although much of the research, as we've come to realize, we want identified, but we really want samples with a lot of health information. And that'll raise the question of how de-identified and how protected our privacy really is.

So what's the real reason for creating biobanks? It's hopefully to study disease and develop therapeutics, but there's also profit-seeking. And so when deciding, as a potential research subject, we need to think about who owns the biobank and what their intentions are. Subjects and researchers should also insist that the consent form clearly articulate who will have access and what type of research that the sample will be used for.

So Carol Ober's study, for example, specifically says that it'll be about women's health surrounding pregnancy, infertility and miscarriage, which in a sense is more of a narrow, disease-specific biobank rather than the KidsGene and TriDOM, the pediatric and medicine ones, which are basically anything about pediatric health or adult health.

Some of the privacy and confidentiality ethical issues can really be solved, in a sense, by technical solutions, with data encryptions and numerous steps in de-identifying. But there was an interesting article that came out two years ago by Kohane and Altman in the New England Journal of Medicine, asked whether genetic samples can ever be completely anonymized in the sense that, you know, this is how we do fingerprinting. When we have a genetic sample, we literally *could* trace people back if we really wanted to. So they actually argue that individuals who were going to participate in genetic research in the future should really view themselves as health information altruists. And they even came up with an idea of having a gene chip in your arm that could be scanned every time you went to an emergency room.

The issue of profit is really important, not just financial profits. Here again let me relate another anecdotal story. This is the case of John Moore, who actually had something called hairy cell leukemia, which, physicians realized, could lead to some very important discoveries, and they kept inviting him back to come back for what they claimed were for clinical follow-up, and really what they were doing were collecting more and more samples, getting bone marrow, getting semen, getting skin samples.

And John Moore, who was from Seattle and getting

health care down in California, was not suspicious until they started putting him up at the Fairmont, and then they asked him to sign a consent form saying that he would ask for no profits. And so he asked a lawyer to look into what was going on, and the lawyer pointed out that a patent had been already been applied for. And he actually went to court, and it went all the way through the California Supreme Court, which basically said that he did not have a right to his genetic material because it had been converted, and so therefore, really was the product of the researchers. And, in response to this, people like Sharon Terry have become much more aggressive in their role as advocates for their genes.

And then the question comes whether the donors will have access. So as I said, Sharon Terry, in a sense, was probably influenced both by the John Moore case and by the Canavan story, which is a story of a researcher, Dr. Greenberg, who got a lot of samples from an Orthodox Jewish community to study Canavan disease. When they finally discovered the gene, they did go and license it, and then they prohibited all the labs that were using it, unless they paid a relatively high licensing fee. And so as I said, many of the individuals from this community were no longer able to afford the test that was specifically designed for prenatal diagnosis for their community.

There's also the issue, as I want to keep saying, that genetic information is probabilistic information. With some diseases, it's easier to understand inheritance than others. You get someone with Canavan's disease, or PXE. You have a Mendelian type of autosomal recessive in heritance. It's easy enough to understand what the inheritance is and how, if you are a carrier, the risks to your children are. There are still difficulties in understanding that genotype doesn't correlate with phenotype and that some people will be more severely affected than others.

But as we start talking about alleles like apoE, which, having apoE4 just increases your risk -- 29-fold over the general population -- doesn't really tell you whether you're going to have Alzheimer's or whether you're going to develop heart disease or any of the other problems that can arise from apoE and its associated diseases. And so the whole issue of how we will get genetic counseling when we already have a shortage of the number of genetic counselors, and we don't even have a way of billing, which of course in this country is critical.

So this issue of who owns the biobank raises a couple of other stories, right? So Sharon Terry has made it quite clear in her story with PXE that she's a part owner of any research that's going to use the samples that she's going to distribute to you. We had a case here again in Chicago this past year. Dr. Catalano vs. Washington University. Dr. Catalano is a very famous prostate disease doctor, and he's the developer of the PSA (prostate specific antigen). And when he moved from Washington University, he actually wrote to all the individuals who had given samples into their biobank and said, "Give me permission to move these with me to Northwestern University." And many, many of the individuals actually signed consent forms and sent it to him.

University of Washington said, "We own these samples. We've been storing it. This has taken our money, our space." And they went to court. And the answer is, at least for now -- we'll see if it goes further -- but Washington University won. And so this whole notion, these biobanks, when you ask, why are we collecting it, they're very valuable. You just have to find the right buyers.

And so we need to be thinking, though, are these national treasures, and should we be treating these at the national level, or should we have Northwestern creating its biobanks and University of Chicago creating its biobanks, and never the two shall meet unless we can come up with some contractual deal which will give University of Chicago first authorship at all moments.

[*Audience laughter.*]

It gets more complicated when we move from the consenting adults into pediatrics, because here, now, we have a third party consenting for another person. This also can happen even with Alzheimer's research when you're dealing with individuals who are no longer able to consent for themselves. There are clear data that show that you are much more likely to consent for your mom who's infirm and for your child to be a research subject than you are for yourself.

So this whole notion -- and I was never surprised at what I would do for dear old mom -- but I was surprised that the data shows that we'll actually be more willing to enroll our own children in research than we would for ourselves. So this does raise some

interesting issues.

For example, in the federal regulations, it's very clear that you should do research on adults first. So the question is, are there reasons to do the research in children? And if there are reasons, do parents have the authority to sign a blank check? Well, I as a competent adult, can decide that I want to donate my samples and you can use it for any purpose. Can I do the same thing in pediatrics? And even if we were to answer those two questions as a yes, we still would have to ask, so what does a parent need to know in order to content, and what do we do when the child reaches majority? Do we go and reconsent? And if we can't reconsent, can we still use the samples or do we have to discard them?

And some would say that there are many things our parents do to us that unless we affirmatively decline as adults -- we maintain our own name, we retain the religion they gave us and many of the other things that they gave us, and so it takes an active role to actually withdraw. And so that there should be permission to continue to use the child's sample, with due diligence, of course, in trying to locate the child to inform him that his sample is still being held and still being possibly used in research.

There are children. There are also, then, identifiable samples, whether it's the Hutterites, whether it's the Ashkenazi Jewish population. And so we get into the question of, are there additional consents that need to be used with the whole issue of group harms and group consent. So group consent might work if we're looking at the Navajo Indian population, but it's going to be much harder in the Ashkenazi Jewish population, where there is no one defined leader. In fact, some of us would say, if you had a hundred Ashkenazi Jews in this room, and I'm one of those, there would be a hundred of us who would think we're the leader of the clan.

[*Audience laughter.*]

So the question is, who would you ask for that group consent? In assessing harms and risks, you have to consider, though, the potential of group harms, which aren't included currently in IRB reviews and aren't even mentioned in the federal regulations. And what does it mean to have group harms? Well, I never participated in any of the research about breast cancer or colon cancer or Tay Sachs disease or Canavan's disease. We have a prenatal Ashkenazi Jewish screen here in Chicago that tests for nine diseases. So just by acknowledging that I'm an Ashkenazi Jewish individual, therefore, life insurance, health insurance and all these other policies should sort of go, "Uh-oh, she's expensive," in a sense.

And so, does it matter that I didn't participate? Because I'm a member of this more isolated community, so that information can be known about me. And this can also occur with other racial and minority ethnic populations. And so you could imagine scenarios where you can have a grandmother agreeing to participate in research and a granddaughter agreeing to participate, but the mom, or the sandwich generation, who's refusing, and yet information will be learned about her, right?

If my mother has Huntington's disease and my daughter has the gene for Huntington's disease, then I, by definition, have the gene for Huntington's disease. Of course, if my mother has the gene and my daughter doesn't, we still don't know anything about me. But there are ways that information will be learned. And so to what extent, as a third party, about whom information may be gained during your research, do I have a right to consent or to refuse to consent to be part of your research, and how will this interfere?

And again, I can go into anecdotes, but the five-minute mark has hit. So I'm just going end and answer questions by just saying that population genetic databases and biobanks offer the potential for great research opportunity, but like all research, they raise serious ethical issues about consent, access, ownership, benefit-sharing, and group and individual benefits and harms, and that we need to integrate human subject protections into research design at all stages of data collection, storage, and usage. Thanks.

[*Applause.*]

Q & A with Dr. Ross

Q: I think that was a wonderful talk, and I agree with all the issues that you raised. In practice, is there a template for moving ahead? Is there a source we can turn to? As affiliates with different studies, we want to create a standard that's appropriate, or at this point...?

Ross: So the NCI Rand report, for example, came up with a really nice set of guidelines for tiered consent.

Now, some of us may not agree to that; we really want only people who will agree to that fourth tier, which is, I'll agree to have my samples stored and used for any purposes. But at least they started giving suggestions for how samples can be collected, and how consent should proceed.

I think there is a whole literature on trying to come up with at least some solutions. And I think some of the solutions are also being pushed on us, right? We weren't expecting the Sharon Terrys of the world. And that's an important statement because it makes us realize that research really has to be a collaborative enterprise; that we can no longer ignore the research subjects or human research participants, whatever name we want to give them. They are playing an active role.

Now, they're not discovering the genes, and yet, if we can't do the research without them, then it's going to have to be a shared process. So it does raise some important issues. I think it's coming in both directions. There is a literature out there, and I think we are getting some standardization about what it means and what type of harms and when do we have to get secondary consents and things of that sort.

Q: I have a follow-up question. One item you didn't touch on is the possibility of giving consent for more general use after the respondent has died. And that doesn't address all of the issues -- probably only about half of them. But it does seem to address, let's say, half the issues. Because that's something that's...

Ross: No, no, no, a really important question. Then, of course, if you're going to be reporting back results, do you need to put in there, if we're going to report back results, do you want us to report these results back to living descendants who may have a real interest in this. So there's a whole set of issues. Just like we have to worry about when the minor reaches majority, we do have to worry about what to do when the individual dies. And so that, too, should be included in...

Q: Not just that one. The group harms are also relevant when we're talking about...

Ross: They don't go away just because you die. Absolutely. Your family is still alive, and they can be very much affected by research that's done on your samples.

Q: Would you say the University of Chicago is selling samples?

Ross: No. We are not. In fact, in all three of our biobanks, it's very clear that we will not sell samples. NuGene was designed so that samples can be sold, and some of the other biobanks are. And actually, give us credit that our policy not to sell samples was written into all three consent forms.

Q: [*Inaudible audience question.*]

Ross: No. I was referring to what is being done at Northwestern University. So at the University of Chicago, you can't sell biobank samples but that doesn't mean that you can't come up with a research project that we wouldn't share the sources with you and collaborate with you in the research process itself. So not to be unfair to those who are going to buy and sell samples, the question is, how much different is it if we say, well, we'll share our raw data with you as long as we're part of the research project, and therefore, also a part of any discoveries that come out of it and things of that sort, versus that you're actually selling it for dollars. How much different is it, because there's profits in both directions.

So I don't want to suggest that clearly, it's wrong to be able to sell samples. The difference is, we're on the South Side of Chicago. We have a very vulnerable population, so I'm actually quite pleased that we're not going to just sell them outright.

Q: So many our studies are collecting specimens that we're storing for future use. And we may have even asked permission, hopefully, from research subjects whether we could do that. I'm guessing that most of us, and I can speak for our own study, didn't anticipate some of the issues you raised, like post-mortem use, reporting to other family members, etc. Do you think that the kind of bioethics principles will evolve to the degree that makes the data we've collected useless, or will there be circumstances under which we can still ethically use these specimens, or...?

Ross: There are lots of different questions in that question. Just because someone dies doesn't mean you have to stop using the samples. The question is, what does it mean if you said you were going to report back? If you didn't specify that you were only going to report back to the individual, then you *could* report

back to family members. But if you said you were only going to report back to that individual, then in a sense, you've actually answered the question, whether you realized you were answering it or not.

So the real issue... So the case that I keep talking about was the Framingham Study, which was done in the 1940s and '50s, which wanted to look for risk factors of cardiovascular disease. Well, back then, nobody was doing genetic studies. We weren't even sure exactly how many chromosomes until the late '50s, right? And so the question then came up, could they use those stored samples for genetic studies?

Well, genes are also risk factors. But the decision at the NIH and the NHLBI was no, that you needed specific consent for genetics. Now, I would say that decision was made in the 1980s when we were going through a period of sort of genetic exceptionalism. We may not have left that period of genetic exceptionalism, but we're starting to realize that genetics isn't that future diary we used to claim it was. Even if I could read out your entire gene map, I don't know who you're going to vote for in November -- although since you live here in Chicago, it's probably more predictive than any genetics you have.

So I think that that was why NHLBI made that decision. And so it wasn't until the more recent Framingham II and Framingham III that specifically allowed for genetic studies. Well, what if we wanted to look at protein mapping? Is that included? Well, now, we're writing consents that are broad enough to allow for new technologies, tandem mass spectrometry and things of that sort. But that wasn't written into the earlier consent forms. Hence the reason for wanting broad and vague consents!

And I think we're writing it into consent forms to cover future research that we haven't even been able to name at this point. Whether that's necessary, it gets to the question of, what are we really trying to protect and for whose benefit are we using these samples. So hopefully we're just going through a phase in that regard.

Q: I'm wondering if you've seen anything in the consent form writing that speaks to respondents who decide after they've given a biological sample that they'd like to be removed. We've seen only a handful, but in other studies, too, where respondents decide they'd like to have their data points removed. And it's such a process for everyone involved to remove those, and I think with biological samples, there's even some other issues with it.

Ross: So again, it should be clarified in the consent process: What does it mean to remove? So does it mean that you're going to remove the data for future use, but that it doesn't stop it from use that's already in progress would be probably the smartest way to write it. Some of the researchers in Europe want to be writing the policy such that if you ask to have your sample destroyed, that it will be anonymized but not discarded, so that they can continue to store and use the samples, albeit in a more limited way.

I don't think that would fly here in the United States. I think we would say that if you have the right to withdraw, it's to withdraw your sample. That doesn't mean you have the right to withdraw your sample to give it to Catalano to bring it from Washington University in St. Louis to Northwestern, but it does mean that you have the right to have it removed from the biobank or the repository so that it won't be used for future studies. And I think that needs to be, again, clarified. And I'm not sure that that's going to be an ethical issue; I think it's going to end up being clarified by the courts.

Gudnason: I may want to comment on a few of these things, because in Iceland, we have gone through a lot of discussions regarding biobanks. And for example, there have been laws passed on biobanks. And there, it says explicitly that a biological sample is not a property, so you cannot sell it, you cannot put it against anything. So it's not...so by law, all these problems are practically taken out.

Ross: Well, many of them... When deCODE got started, it wasn't that simple, right? There's a whole scandal that gets involved with the deCODE and with individuals and with the Icelandic Medical Association being against the whole movement to having this national biobank and trying to discourage patient subjects from enrolling and to say no. At this point...

Gudnason: I want to contest this, because this is not true, that there is a huge, centralized health care database that is full of biological samples without informed consent. That's not true.

Ross: No, the original deCODE arrangement was that you didn't have to give informed consent, but that you

could ask to have your samples removed.

Gudnason: No, that's not true.

Ross: Then it went through an entire community consultation process, and yet you still collect samples from children and the children can then remove their samples when they become adults, but they can only do so if they know. So there's been a lot of changes in deCODE over the years, and it's ethically better, but still…

Gudnason: Well, I happen to come from this country, and I know that this is not how it is. They have never been able to collect samples, biological samples, without informed consent of collecting the samples. What people are mixing it up with is that they passed a law allowing them to mine, practically, the health data system. But that never...

Ross: Fair enough. Right. So it wasn't about the sample, but it was a good mining expedition. And as people consented to the... I think, though, that Iceland is an example of how things should be done; if you go ahead four or five years. Because that became, then, a whole national debate about who should have access and how the samples should be collected and whose consent is needed and what should be done if people want to withdraw.

So that eventually the practice that has come into place in Iceland is acceptable to the population, but I think there was a lot of confusion at the beginning. The fact that it could be done by law in an entire state also goes to the homogeneity of a country. I think it would be much harder to pass that type of legislation locally in this country, let alone at a national level.

Gudnason: But there are a few principles. For example, we are, by law, requested to put in any informed consent, which says that you can withdraw your sample at any stage and refuse the use of your data for any further analysis. But it's also explicitly stated in the law that what has already been done and put into analysis is not going to be removed from your database.

Ross: Right. And that's what I was suggesting as the right solution anyway, but that the samples should be fully removed and it shouldn't be that it just becomes anonymized, which is being suggested in some of the other European countries.

Well, I have a signal from the chair, so let me conclude by saying thanks again for organizing such a stimulating conference.
[*Applause.*]

Panel Discussion: ***Exploring Age-Old Questions: Environmental Effects, Cognitive Function, and Alzheimer's Disease***

Hugh Hendrie, MB, ChB, DSc (Indiana University)

Thanks, Marcus. I'd like to thank the organizers of this very, very interesting meeting. It is a pleasure to be here, a pleasure to listen to the other speakers, and a pleasure to drink all that wine last night.

There's much more emphasis in this conference on methods than what usually occurs at other conferences. In this talk I will present results from the Indianapolis/Ibadan study, the manner in which we've incorporated biomeasures into our design, and the interactions between these biomeasures, genotypes and cognitive outcomes. However, the results presented in each one of these slides raises a whole series of methodological problems that I'll try to mention them in passing.

INDIANA UNIVERSITY

Age-Specific Annual Incident Rate of Dementia and Alzheimer's Disease, Adjusted for Mortality

	Age Group, y	Yoruba Rate % (95% CI)	African Americans Rate % (95% CI)
Dementia	65-74	0.45 (0.30-0.60)	1.74 (0.15-3.32)
	75-84	1.69 (1.29-2.10)	4.29 (2.52-6.06)
	≥ 85	5.71 (4.19-7.22)	9.12 (5.97-12.28)
	Age-standardized overall rate	1.35 (1.13-1.56)	3.24 (2.11-4.38)
Alzheimer's Disease	65-74	1.38 (0.24-0.52)	1.38 (0-2.99)
	75-84	1.41 (1.04-1.77)	3.29 (1.56-5.01)
	≥ 85	5.02 (3.62-6.42)	7.07 (4.54-9.61)
	Age-standardized overall rate	1.15 (0.96-1.35)	2.52 (1.40-3.64)

IU Center for Aging Research

The Indianapolis/Ibadan study has been continually supported by the NIA now for the past fifteen years, and it is refunded until 2011. Dr Kathleen Hall described the study briefly, yesterday. It's a comparative study of the incidence and prevalence of dementia and its risk factors in two community dwelling populations – elderly African-Americans living in Indianapolis, and elderly Yoruba, living in the City of Ibadan, in Nigeria.

The development of the study involved overcoming major practical and methodolodgical issues. For example, recruitment and retention is a huge problem in most longitudinal studies. Our group led by Dr Hall did a wonderful job in community preparation and contacting the community leaders that were necessary to allow us to recruit and retain our study populations. This was exemplified by the fact that in the first stage of our study -- in 1992 -- we had only eleven percent refusal rates from African Americans in Indianapolis, a very low rate, and in Ibadan, it was even lower than that.

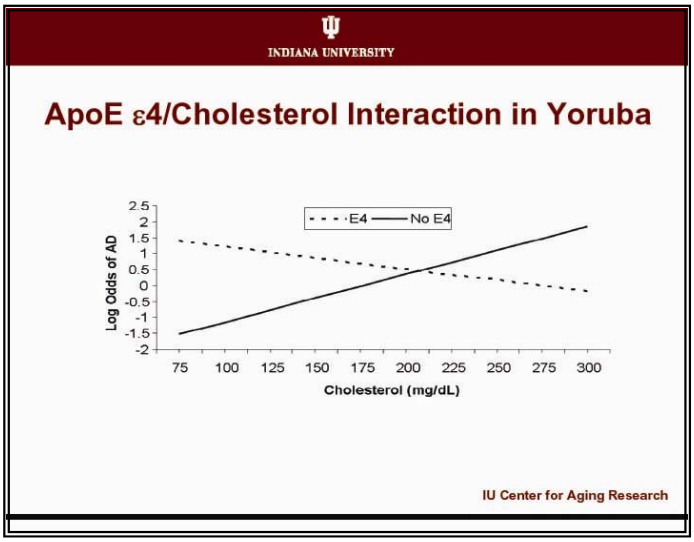
We have developed a screening instrument starting with our studies of elderly Cree in Manitoba. It was designed to provide culture-fair assessments of cognition and function that could be applied to people who were both literate and non-literate. We've now used it in the Cree, African-Americans, the Yoruba, and populations in China and Kenya. Martin Prince is now using it in his studies. It appears to be a very robust instrument in all these settings.

The study design incorporates two stages a screening phase followed by a clinical stage. Based on the results of the screening instrument, the population is divided into good, intermediate, and poor performing. A clinical assessment is then performed on all individuals in the poor performance category, and a sampling is taken from the other performance categories. The clinical assessment is elaborate and identical for both sites.

Standard neuropsychological tests are used with normative values available from both sites. In addition, an informant interview is conducted, as well as, physical and neurological examinations. CT scans based upon clinical necessity have been conducted at both sites in a number of patients, and an additional 50 MRI scans were conducted in African Americans as part of an add on pilot study.

The diagnostic process is elaborate. A consensus conference is performed separately at each site, which both physicians and nurses attend. Then the data (and the faculty) are transported from one site to the other. The data is then evaluated blindly by faculty from the other site, and independent diagnoses are made. Finally, a joint consensus conference with all of the faculty from both sites is conducted. The diagnostic agreements between sites have been good. In some of our earlier waves, in order to ensure diagnostic

consistency, each of the sites would select three or four patients that represented some particular problems, and the opposite site faculty would go out to the houses of those patients and independently assess them.



Our study started in 1992. A prevalence wave was conducted in '92-'93, with incidence waves conducted subsequently every two or three years. Data collection is completed, for the 2004 wave, and the screening process has been completed for 2007.

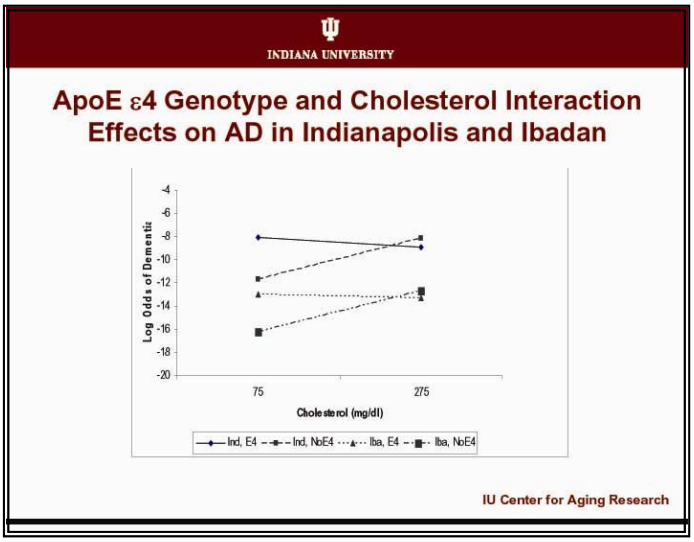
The risk factor model we follow was originally proposed by Dr. Richard Cooper for hypertension in studies involving populations in Nigeria, Jamaica, and Chicago. In order to understand population phenotypic variation, he proposes that the frequency of putative genes in the population be identified as well as the frequency of environmental risk factors, and, most importantly, study how these genetic and environmental factors interact with each other. Genetic environmental interactions will, in my opinion, be the most important focus for epidemiological studies in the future, although this likely will likely be very complicated, as I will demonstrate.

I should tell you that this was not our model when we started. We originally conceived the study as one involving migrant populations, as African Americans largely came from West Africa as part of the slave trade, and now like in two very different environments. Thus, variation in rates would likely reflect the environmental differences between living in the developing and developed world.

Then, in the late 1980's early 1990's two major research developments occurred, specifically in 1989-1990 by two sets of studies. Researchers at Duke identified the possession of apoE4 as a major risk factor for Alzheimer's disease. At the same time, studies involving population genetics exploded. It became impossible to ignore genetic factors in population studies of Alzheimer Disease, hence the usefulness of the Cooper model. This also emphasizes the necessity for longitudinal studies to be adaptable to incorporate new research findings.

The findings from population genetics provided additional benefits for our study. They suggested that all humans came from Africa relatively recently -- 200,000 years ago or less. And, furthermore, if you really want to study complex illnesses that affect human beings, you really need to study them in the homeland of the human race: Africa.
So this is Ibadan: a huge city of over a million people, about a hundred miles inland from Lagos. The Idikan Wards where our study is conducted is in the central part of the city. So, the population is city dwelling, but it does retain some of the characteristics of a village, circumscribed as it is by markets.

Most of the older people in our study describe themselves as traders, all their lives and, by the way, continue to be traders at least until they begin to lose



their function. So there's not this abrupt separation of retirement and non-retirement in Yoruba as in developed countries.

The residents of the Idikan Ward are not wealthy. Most of the families live in these large houses, which contain large extended families. Eighty percent of

them are Muslim.

The African Americans in our study live in Indianapolis, in individual housing with single families, or in senior apartments with other old people. Many live alone. Many *demented* people live alone, leaving us to wonder sometimes how they manage.

So let me review first our already published major findings.

Here are the age-standardized 5 year incidence rates for dementia and Alzheimer's disease for both populations. The incidence rates for dementia and Alzheimer disease were much higher for African Americans than for Yoruba. African Americans were developing Alzheimer's disease at about two and a half times the rate of the Yoruba, but as you can see the association of incidence rates with age were about the same for the two communities. In both sites, the preponderance of cases of dementia was due to Alzheimer Disease.
The other major finding of our study was the difference in association of the possession of APOE 4 allele and AD risk in the two communities. APOEe4 was a major risk factor for Ad for African Americans but not for Yoruba. Similarly, e4 was a risk for increased mortality as well as cognitive decline in African Americans but not in Yoruba.

We have also conducted a pilot study in Kenya. There is insufficient data to arrive at a definitive conclusion from that study, but again there didn't seem to be any relationship between possession of the E4 allele and Alzheimer's disease. So whether or not lack of association is an African-wide phenomenon is uncertain. If it is, it would be unique in world populations so far studied.

Yoruba not only had lower rates of dementia but also lower rates of cardiovascular diseases and risk factors such as diabetes, hypertension, stroke and hypercholesterolemia. Their mean BMI was also considerably lower than African Americans. This led us to consider the possibility that there may be a link between cardiovascular disease and cardiovascular risk factors and Alzheimer's disease. We weren't alone in considering this possibility. There's now a growing consensus among the scientific community of Alzheimer's disease that there *is* indeed some link between the heart and the brain.

One of the first genetic/biomeasure interaction analyses we conducted involved e4, cholesterol levels and risk of AD. There was a significant interaction with African Americans. African Americans who did not posses the e4 allele had increased risk for being diagnosed AD with increasing levels of cholesterol, but this association was not significant for individuals with e4. But this analysis was cross sectional on a limited number of subjects. It did lead us to conclude that the effects of biomeasures such as cholesterol can be modified by genetic factors. So, in order to understand cardiovascular risk more comprehensively, in 2001 we decided to incorporate biomarkers for cardiovascular risk into our study.

We have collected about 1,500 samples from Indianapolis and 1,200 samples from Ibadan. This slide shows the demographics and clinical characteristics of the individuals with biomarkers for the two populations. There were some surprises. For example, systolic blood pressure was higher in Yoruba than we anticipated. But remember, most of the Indianapolis individuals are on anti-hypertensive medications. Co-morbidities were higher in Indianapolis. This doesn't mean necessarily that Yoruba are healthier than African Americans. It's not that the Yoruba are a healthy population; they're not. It may be that the kinds of illnesses we included within the co-morbidity index overemphasized things like hypertension, diabetes, and so on.

When we compare biomarker levels between the populations there are some anticipated findings as well as some surprises. All of these measurements are site adjusted for age, gender, BMI, drinking, smoking, etc.

In the "no surprise" data were lipid measurements. African Americans had higher levels of cholesterol LDL's and triglycerides than Yoruba did. Somewhat of a surprise was that HDLs were higher in African Americans, It is noteworthy however that within African Americans, cholesterol levels are much lower now than in 1999. In 1999, a miniscule numbers of African Americans were on **statins**. In 2001, a quarter of the African Americans were on statins.

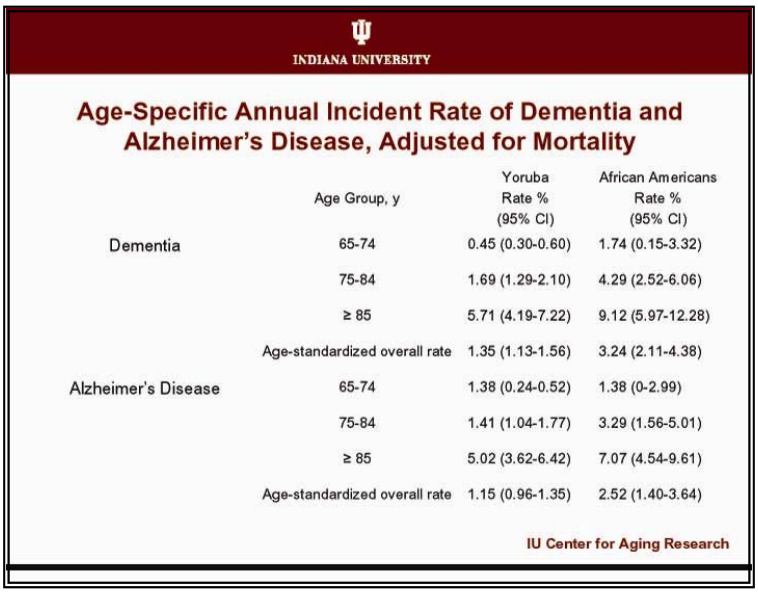
Again, comparisons of measurements of glucose-insulin produced no surprises. Again, you would expect that levels of glucose and insulin were higher in African Americans than Yoruba, even when taking into account the differences in fasting measurements. It was no surprise, therefore, that metabolic syndrome

was much more commonly found in African Americans than Yoruba.

There were some surprises however, in that the C-reactive protein levels, a measure for inflammation, were about the same in both sites. Some measurements of endothelial dysfunction biomarkers were actually higher in Yoruba, and yet the rates of heart disease are lower.

Another surprise: the biomarker for oxidative stress – 8-isoprostane -- was much, much higher in Yoruba than in African Americans, and we're not sure why. It may be dietary related, as 8-isoprostane levels are related to diet. Apparently, consuming potatoes leads to higher isoprostane levels. (Bad news for a Scots Irishman.) Yoruba eat lots of tuberous yams, so this may account for the high levels.

When we analyzed the new biomarker information, we again found an interaction with lipid levels, APOE and AD diagnoses for Yoruba very similar to the ones we already reported for African Americans. And remember, E4 is not a risk factor in Yoruba.


INDIANA UNIVERSITY

Age-Specific Annual Incident Rate of Dementia and Alzheimer's Disease, Adjusted for Mortality

	Age Group, y	Yoruba Rate % (95% CI)	African Americans Rate % (95% CI)
Dementia	65-74	0.45 (0.30-0.60)	1.74 (0.15-3.32)
	75-84	1.69 (1.29-2.10)	4.29 (2.52-6.06)
	≥ 85	5.71 (4.19-7.22)	9.12 (5.97-12.28)
	Age-standardized overall rate	1.35 (1.13-1.56)	3.24 (2.11-4.38)
Alzheimer's Disease	65-74	1.38 (0.24-0.52)	1.38 (0-2.99)
	75-84	1.41 (1.04-1.77)	3.29 (1.56-5.01)
	≥ 85	5.02 (3.62-6.42)	7.07 (4.54-9.61)
	Age-standardized overall rate	1.15 (0.96-1.35)	2.52 (1.40-3.64)

IU Center for Aging Research

What happened to the interaction between lipids and e4 in African Americans? Well, the problem with African Americans now, is you have to enter statin use into any model. So, now the complexity increases. Models have to include not only biomarkers, but treatments which influence these levels. Our statin analyses were difficult to interpret. While individuals who took statins at baseline were at lower risk for developing cognitive decline, individuals who had stopped taking the statins were at the lowest risk. By the way, the effects of the statins did not appear to be mediated through their interactions with lipids or CRP. Taking statins and then stop taking them will make you better.

The remainder of the slides shows a very, very preliminary analyses of biomeasures and cognitive decline over a 3 year period. They should be observed with a degree of caution.

So, in African Americans, for instance, higher levels of insulin were associated with more cognitive decline. There were also many interactions between levels of biomeasures: APOEe4 and cognitive decline, for example. Illustrating again the necessity of taking into consideration, at the same time, genetic and biomarker information.

I would like to finish by paying tribute to our wonderful group of investigators at Indiana University and the University of Ibadan. This is truly a collaborative endeavor,

And finally, a tribute to our old friend and colleague, Ben Osuntoken, without whom this project would have been impossible. Ben, alas, died before he had a chance to see it come to fruition.

[*Applause.*]

Q & A with Dr. Hendrie

Q: Thank you for that very interesting talk. How do you account for differences in selection pressure, so that maybe, by the time you're studying elderly in a difficult-to-live-in environment, they're selected to be much healthier than elderly who can survive in an environment with many more resources?

Hendrie: I think that's good question, which we don't know how to answer. In a way, selection doesn't matter to us. What we're looking at is two populations that have survived, and our eventual goal is to bring the explanation of the differences of the populations to the molecular level. So, if we can find out, molecularly, why's there's differences in these populations, selection doesn't matter terrifically to us.

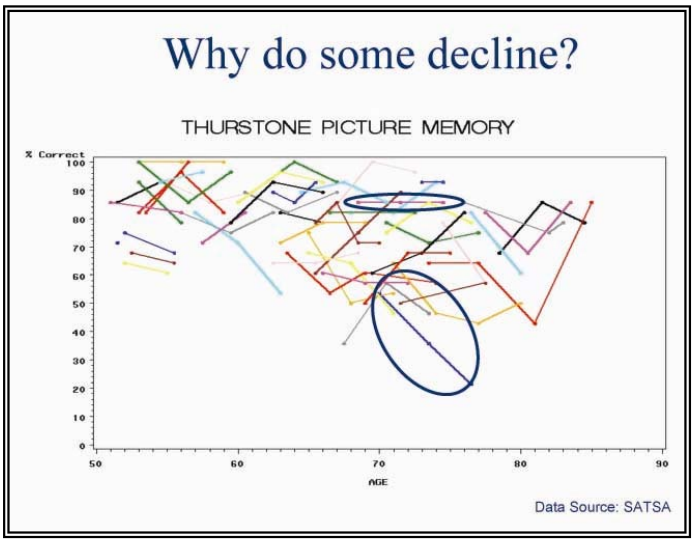
The selection theory would have to say that something happens in early life that is related to Alzheimer's disease, which, later in life, makes people more resistant to it. And I'm not sure that I know any

selection theory that says that. People who die in Nigeria, for instance, die of inflammation, infections, pneumonia, and so on.

The other interesting thing about it is that there is no difference in mortality rates between the elderly Yoruba and the elderly African Americans. Actually, the elderly Yoruba are still dying at higher rates than the African Americans. But within that mortality rate, the effect of dementia is exactly the same in the two populations, so the added mortality risk for having a diagnosis of dementia is almost identical between African Americans and Yoruba. So, something is happening, which is nice in a way because it's an external validity of our diagnosis of dementia, but it's also surprising in a way. It doesn't say much about our first-world health contribution to the care of the demented patient if the result is still no different than in Yoruba. So that's a hard question to answer.

Chandra Reynolds, PhD (UC Riverside)

Thank you. I want to thank the conference organizers for inviting me to speak today. I'm just doing a little equipment change here.

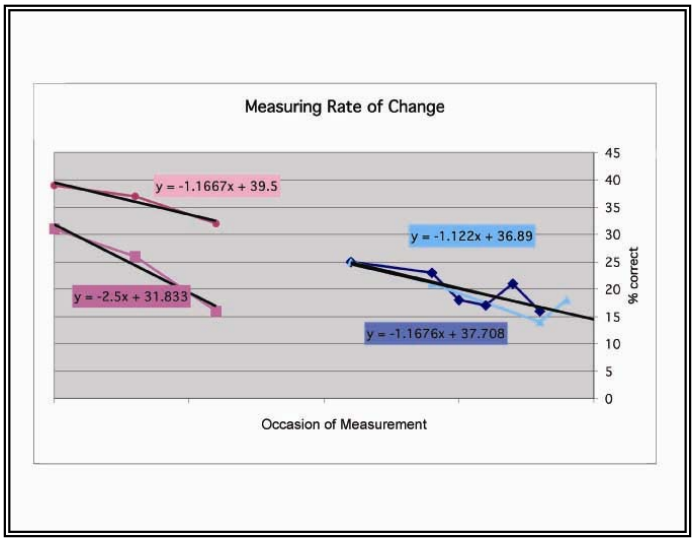


So today, I'm going to share with you some of our models of cognitive change in the SATSA study -- which is the Swedish Adoption Twin Study of Aging -- and I mentioned a little about the study yesterday. I'm going to talk, in general, about genetic and environmental contributions to cognitive change, and then, how we've incorporated measured candidate genes, as well as considered gene-environment interactions.

So one of the questions in our quantitative models is, why do some show decline while others show stability? And what this plot is is the Thurstone Picture Memory Task, which is an episodic figural recognition task. On the Y-axis, you have percentage correct out of 28 items given, and along the X-axis, you have age of the participants. And this is just a random selection of twins from the SATSA study. So you can see for some individuals, they're performing poorly over age, while others maintain stability across age, even at the same age range. If you were to look across the entire sample, though, what you *do* see is relative stability of the sample prior to age 65 or so, and then afterward, you start to see a fanning effect, where you see some individuals showing decline while others maintain some stability. And I might add that I've not included demented individuals in this slide.

Now, in setting out to measure a rate of change, we've considered multi-level models, and for simplicity, I'll just give the general idea here. We have two pairs of twins here -- the ones in the upper-left are a pair of DZ twins, and the pair in the lower-right -- essentially, the curves overlap -- is a pair of MZ twins. And so the idea is to characterize cognitive change into a set of parameters that doesn't rely, necessarily, on the total number of points, but is trying to reduce the number of parameters into the essential mathematical equation that you could characterize the change as a curve. This is just a linear model, here. For the DZ twins, you can see that the better-performing twin has a rate of change per year of about one point of loss per year, whereas their co-twin has a more than double -- about two-and-a-half points of loss per year -- rate of change. For the MZ pairs, they essentially have no difference in their rate of change, and that's what we're looking for -- not only the lift of the curve, so to speak, of the trajectories, but also the rate of change.

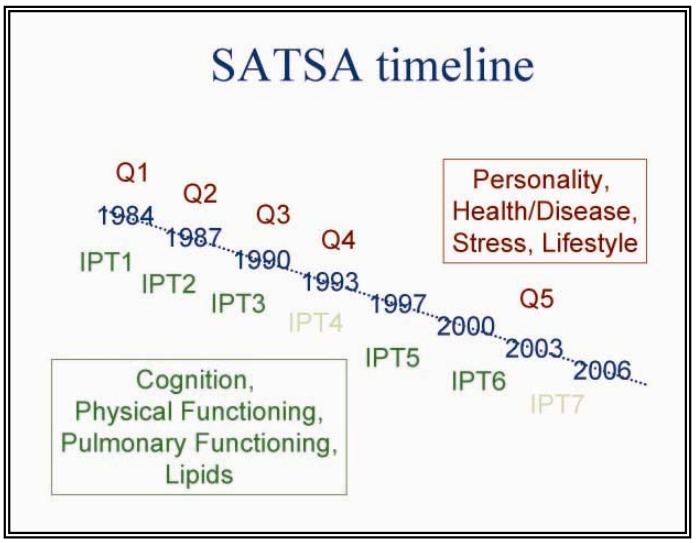
Now in the SATSA study, one unique aspect of it is that we have a sample of twins reared apart, and a sample of twins reared together. It's a population-based sample from the Swedish Twin Registry. It's multi-disciplinary, and for the purposes of the cognitive battery, we have, now, six longitudinal in-person occasions that I'll discuss. We also have a seventh wave that is now under collection. These were



done in three-year rolling intervals, which will become clear on the next slide. And we had one seven-year gap in funding, in which we just essentially performed the telephone screening.

In this slide, you can see that the SATSA study started

in 1984, and was primarily funded by the NIA. In fact, most of these occasions were primarily funded by the NIA, as well as some Swedish sources. So in 1984, we had the first questionnaire wave, which was sent to the entire SATSA sample. And in the interim, we had

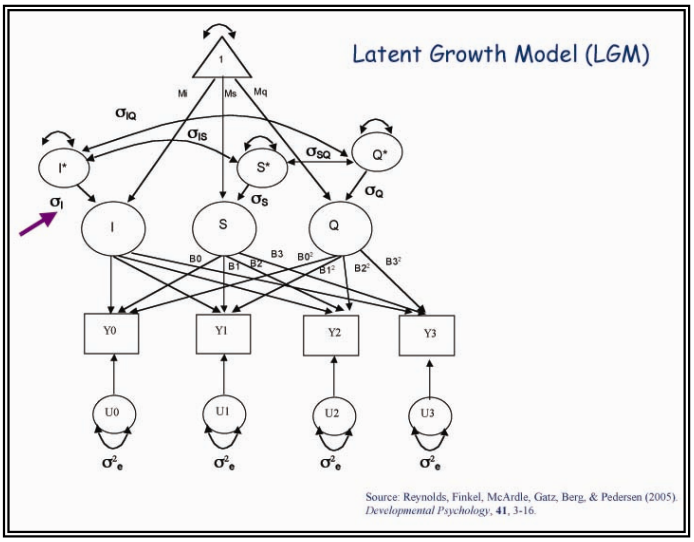


what are called in-person testing waves, where we invited twins 50 years and older to participate. The in-person waves primarily focused on cognition; physical functioning, including handgrip and other adaptive functioning tests; pulmonary function; and a blood draw for lipid and hemostatic parameters. The questionnaire phase included some personality measures, some self-report health and disease, stress in terms of life events, and lifestyle factors. And so, again, we had rolling three-year intervals for the in-person testing phase. At IPT4, there was no in-person testing *per se*, but a telephone call to check their mental status and keep in touch. We then regained some funding, and so we conducted IPT5 and IPT6. And we're now in the midst of finishing up the in-person testing occasion seven (IPT7).

So for the psychometric cognitive battery, we have essentially four primary domains. We also do mental status exams -- the MMSE -- which I haven't listed on this slide. But for our quantitative models, we've considered several tests in the verbal domain, spatial domain, memory, and speed. So our key tests, that I'm going to talk at some length about, are the Information subtest (the Swedish Wave Information) and the memory tests because they show some interesting patterns of genetic and environmental influence over age. And I also just want to mention the lipid panel that we have at most waves. We just have the HDL baseline. But I will talk a little bit later about what our future plans are with lipids and cognitive change, and I'll have some preliminary analyses of that.

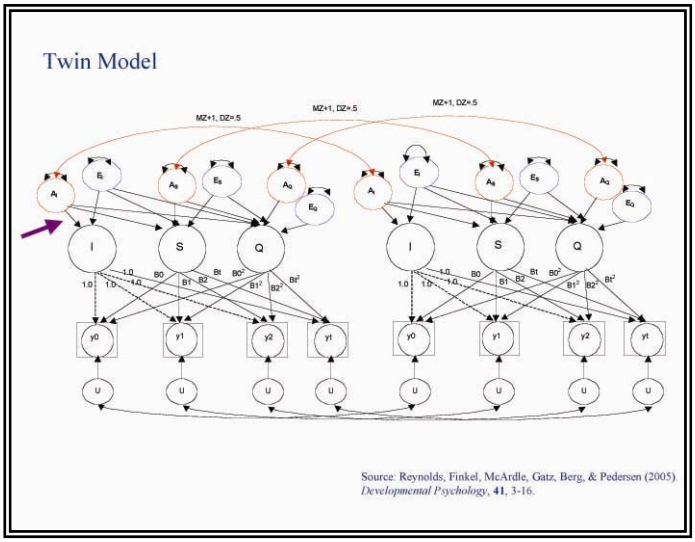
For the quantitative analyses of cognitive change, in 2005, we published a paper which included 798 non-demented twins with cognitive data across thirteen years. So, this included up to IPT5 and 362 pairs, most of which had at least two time points of data. We included any time points of data that they had, from one to four time-points. Again, they had to be 50 years and older at the first testing occasion, and our average age at the first testing was 65 years, and we had primarily females.

So this is a structural equation modeling representation of essentially a multilevel model that we fit to the cognitive data. So, what you have here in the squares are observed data from Y0, Y1, Y2, Y3 -- these would be a set of cognitive scores over time. So, these could be, say, the WAIS Information score at time-one, time-two, time-three, time-four, except that we've called time-one the baseline. Now we've characterized the curves in the sample in a non-linear fashion, because for most of our measures, a non-linear model is more appropriate than a linear model. In this model, you have a latent intercept, which describes the lift of a curve on the Y-axis, a linear slope, which in the context of a quadratic model depends on the centering of the age-predictor. We centered our age-predictor on age 65, so this represents, essentially, the tipping-point of the curve at age 65. And then, we have a



quadratic term, which is reflecting the acceleration in the curve over age. Now, this model estimates -- and this is using a convention of Jack McArdle's -- where a triangle represents inclusion of the means into the model -- we can look at the sample mean intercept, mean slope, and mean quadratic term. And then, of

interest for individual differences, we'd like to know how individuals differ from those average growth parameters. So, we're really interested in these sigmas here, which are reflecting, essentially, the deviations of

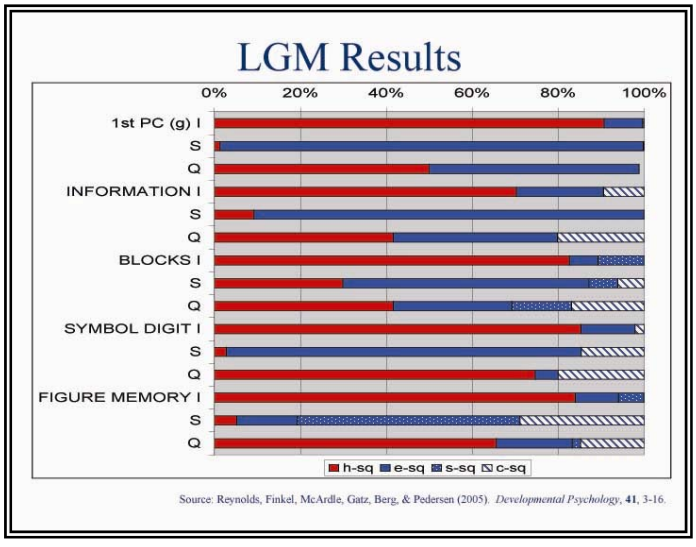


individuals' own growth curves from the sample curve. And of course, because a growth model doesn't necessarily represent all the variation in scores that you might see in a trajectory, these U's here represent the unexplained variation, or specific measurement occasion variation not attributable to the growth curve. Now this model, here, is just for a single individual. But in the context of a twin design, we'd really like to consider the genetic and environmental variants that make up this total variation. And so these latent growth parameters become our phenotypes of interest, and we want to decompose the variation in those phenotypes into genetic and environmental sources.

So in the next slide, we have essentially two of these models -- one for twin A and one for twin B. And you'll notice there has been some shift in the model, where we've taken those sigmas and decomposed them into additive genetic components; and for simplicity, just non-shared environment. With this data, we do also, because of the rearing status, decompose the variation into shared rearing environmental variance, and then correlated environmental variance. But these are not shown on this slide. So, for the additive genetic variance, we have a factor that impacts intercept slope and quadratic – something that just impacts slope and quadratic, and then the unique genetic
factors on the quadratic not explained by the prior two factors. Across the twins, you'll notice that there is a double-headed arrow, which indicates that these latent additive genetic factors are correlated depending on the twin type. So identical twins are going to be correlated perfectly for genetic factors, because they share 100% of their genes. So this correlation here is fixed to 1 if you're an MZ pair. It's fixed to half if you're a DZ pair because, on average, 50% of the segregating genes are shared between fraternal twins.

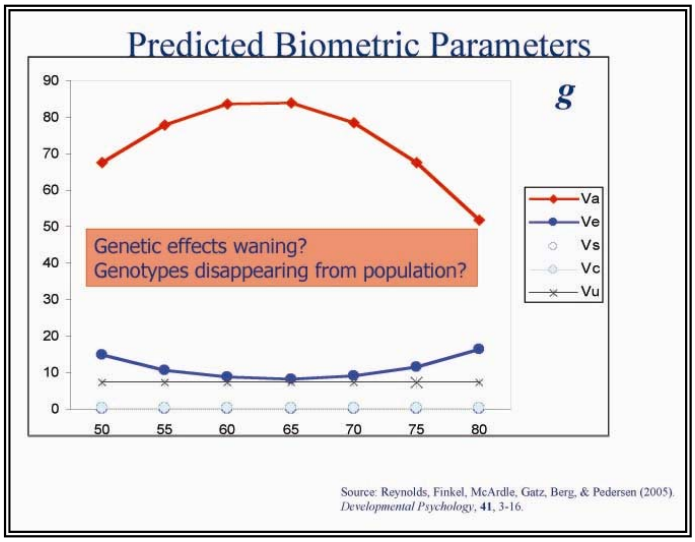
And then here, we allow some correlation for specific occasion measurement variance, but we do not decompose it into genetic and environmental variance. When we look at the genetic and environmental variance associated with these growth parameters as outcomes, what we've learned here is that we see significant and high genetic contributions to variation in the intercept. So genetic variance, here, is coded as the red portion of the bar. And the blue is the environmental variance, which...the filled-in blue bar is the non-shared environment, which is the predominant source of environment.

So we see greater than 80% heritability for the first principle component of all our cognitive tasks. And the others are somewhat less than that, but still high. If you look at the genetic variance for slope, on the other hand, you'll notice that the red portion of the bar is much smaller. So, for the linear rate of change at age 65 in this model, we see negligible to small heritabilities, and predominantly non-shared



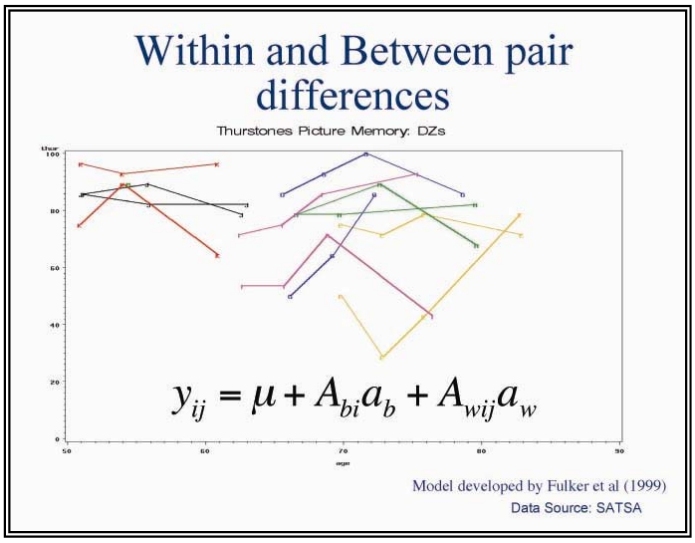
environment, describing the tipping of the curves. This was meant to show you the variation for the quadratic component, which is somewhat in-between, so we'll have to skip over that...which is somewhat in-between the intercept and slope, we see moderate heritability for the quadratic term, in comparison to linear change. And so, we've concluded, based on our

findings, that although linear change is not very



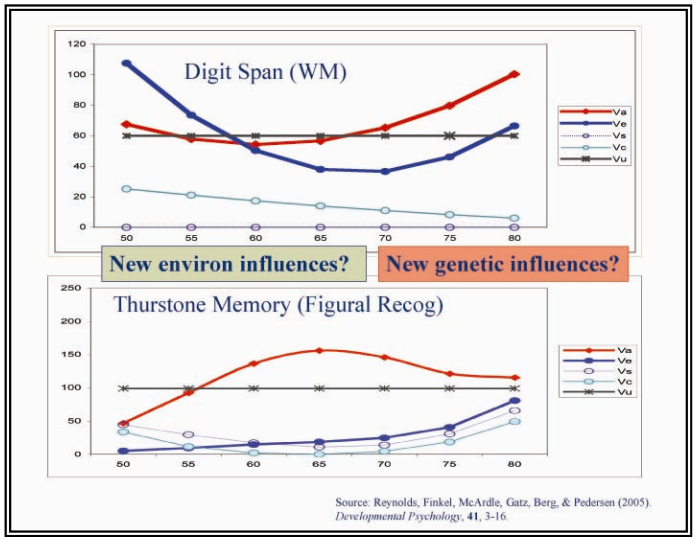
heritable, quadratic change is more heritable, and is more likely to be correlated with the intercept than is linear change.

We constructed expected variance components across age, rather than the heritability of the individual growth parameters, because to some extent, that is dependent on the research design. We chose to center the growth model at age 65, and so you have particular interpretations of the intercept and slope terms, but if you calculate expected variance components across age, based on the growth parameter results, you'll see the curve of genetic variance is decreasing after age 65. And on the other hand, non-shared environmental effects after age 65 are increasing with age. So we could perhaps suggest, then, that genetic effects may be waning, or perhaps some mortality, or selection



effects, where you have genotypes disappearing from the population. This was true for our first principle component -- or our measure of G -- and it was also true of most of the spatial, and speed, and verbal measures. So you see decreasing genetic variance, and increasing non-shared environmental variance. With the memory measures, and including the Information subtest, which one could suggest has aspects of semantic memory processes, what you find is that the red line, which is indicating genetic variance, is increasing after age 65, as is non-shared environment.

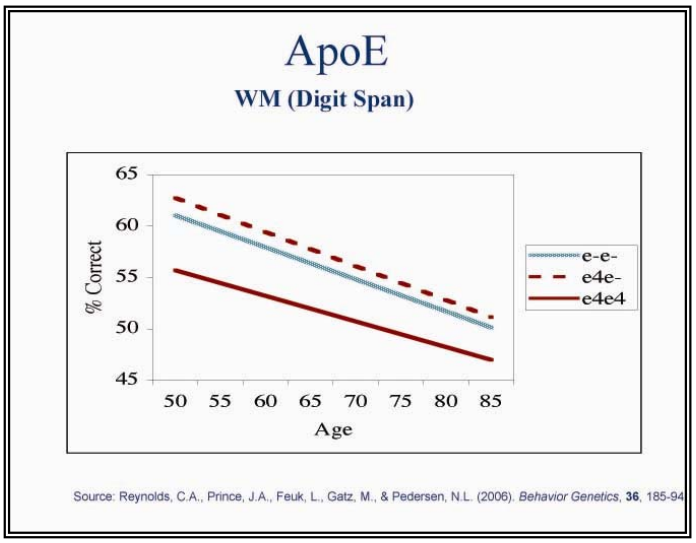
This is Digit-Span up at the top, and at the bottom, we have Thurstone Memory, which is the episodic figural recognition task. Here, again, you see this increasing genetic variance, and increasing non-shared environmental influence. So this may suggest some new genetic influence is coming on-line, or new environmental influences that are accumulating. We



can also suggest that maybe these increasing environmental influences may be actually attributable to gene-environment interaction. In our simple models, we usually calculate the independent effects of genes and environment, but not their interaction or their correlation in our typical twin models. So this increasing non-shared environment may signal some gene by environment interactions that we could pursue further.

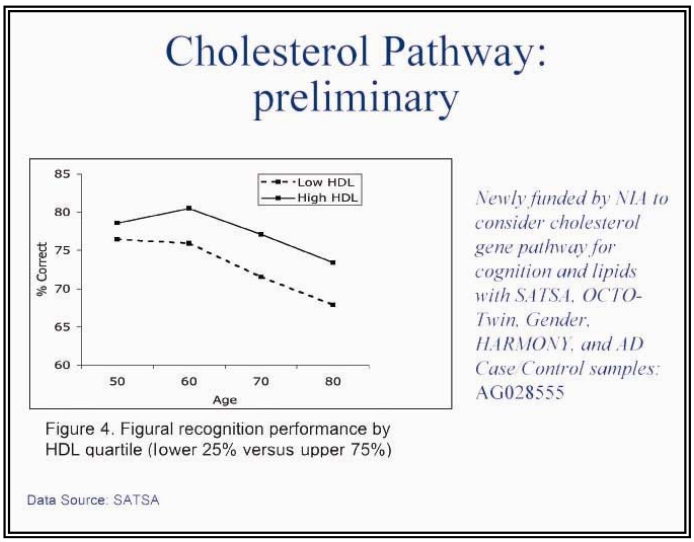
First, I'm going to talk a little bit about including some gene candidates into these growth models. Particularly for our memory measures, we're interested in some associations with, perhaps, gene candidates in the amyloid or cholesterol pathways, such as APOE e4, or candidates in learning and memory pathways, which show a down regulation with age, such as the serotonin-2A receptor. So first, I'm going to describe

the model that we use for looking at gene candidates using twin data.



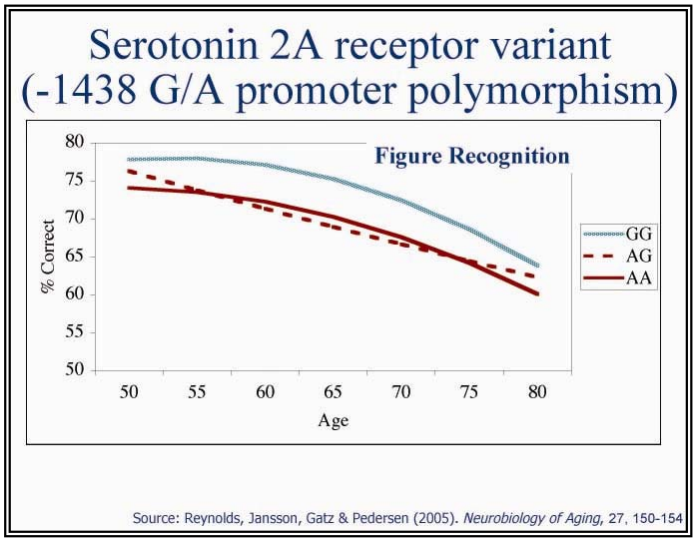
And what I have here on this slide are pairs of DZ twins, or fraternal twins, and they're color-coded. So this is the 'red' pair in the upper-left, this is one pair -- right here -- with one twin showing stability on this picture memory recognition task, and the other twin showing some decline. So, we'd like to know whether or not these differences between pairs are due to differences in genotype. So, the model that we use is to, say, take our outcome, which could be any one of our growth parameters -- intercept, slope, or quadratic -- and it's a function of *between* pair-differences, which are described here, and *within* pair-differences.

And we also fit this model to identical twins, which, of course, never differ on genotype, but which can help us with background variation estimation. So, this is



what we find with ApoE4. We have association with both intercept and with rate of change in Digit Span; we didn't need to fit a quadratic model -- it's the only cognitive task we could fit a linear model to. So you have the double-E4 expected trajectory here -- lower performance across age, and less steep rate of decline. What's interesting about our data, and which is somewhat conflicting with what would be expected, is you actually see... And here, we've included just E2, E4, and E3-4. You see equal performance between the non-E4 carriers and with the single E4 carriers in this growth model.

And here's another set of results with our Serotonin-2A receptor, and this is with the Thurstone Picture Recognition task. Those who were carrying the A-allele, either heterozygote, or the double-A homozygote, show worse performance than those that are double-homozygote for the G-allele.



And I just want to show you some preliminary findings with HDL and cognitive change, not with the gene candidates yet; but just to let you know that we've just been, two weeks ago, newly funded by NIA to consider the cholesterol gene pathway for cognition and lipids, and we're going to be considering some dynamic models of change. We have the lipids at the same time-points that we have the cognitive data. So we're going to be considering some dynamic relations across time, and including some gene candidates in the cholesterol pathway. And I might add that we're including multiple twin studies. It'll be more than just SATSA. OCTO-Twin, Gender, HARMONY, which is the Study of Dementia in Swedish Twins, and an AD case-control study.

I'm going to end with some of our considerations of

testing for gene-environment interaction, and here we use the identical twins in our analyses, because the only reason that they should differ in terms of their cognitive trajectories is due to environmental reasons, as they do share their genetic factors. So what we've done is consider the growth curve, parameter estimates in MZ twins, and I'm going to show you results primarily for the Information quadratic component. And what this plot is, is the intrapair differences in their quadratic parameter for growth. And on the end, here, we have identical twin pairs who are E4-absent, their differences being in the quadratic parameter, and those who are E4-positive. And so what you see is much less difference in their quadratic, or their acceleration in cognitive change, if the identical twin pairs carry an E4 allele than if they don't. There's much more variation in their quadratic change. And we've discovered some findings with the serotonin transporter gene, as well. What we've done to explore this further, is what could be some of the environmental factors that might interact with one's genotype to consider measured environments. The findings on this slide would suggest that those who are not carriers may be more sensitive to environmental factors than those who are carriers of risk alleles.

So what I have here is a plot of the pair differences in the Information quadratic term and the pair differences in the CESD depression measure. And so, what you find is that for APOE e4 carriers there's no association between differences in depressive symptoms and differences in quadratic change. But you do find a positive association in the non-E4 carriers, such that those who don't carry an E4 allele, you find that the greater the pair differences in depressive symptoms, the greater the pair differences in quadratic change. We have similar findings with an estrogen receptor allele, which I will just mention in passing.

So what have we learned from the SATSA data with respect to cognitive change? We find that the environment increases after age 65, and that would be supportive of stochastic theories of aging, or accumulation of environmental factors, or perhaps genotype-environment interaction.

Genetic variance decreases after age 65 for non-memory-loaded measures; and perhaps some new genetic effects for memory traits are coming on-line. We've considered candidate genes, such as APOE e4 and serotonin-2A, and found associations with certain memory traits. We *do* find some genotype-environment interaction evidence for semantic memory change.

In the future, we'd like to consider further distinguishing cognitive decline from dementia, and from terminal decline. The results I showed you today had excluded demented twins. We're more concerned now with including them, in ways that make sense, perhaps doing combined growth-survival model analyses.

And finally, I'd like to end with acknowledgments of both my collaborators on these projects, and the NIA and other sources of grant funding. Thank you.

Q & A with Dr. Reynolds and Dr. Hendrie

Q: The [Gini] Index and other measures of inequality are quite low within Sweden. And so, how do you think your results might look different if you were to look in a population where there's more inequality, more heterogeneity in the environments that people are experiencing -- for instance, the partitioning of variance between environmental and genetic influences, the nature of the gene-environment interactions or their magnitude? Just wondering what you think about that.

Reynolds: Right. I actually think there would probably be more total phenotypic variation...greater variation in total. And I think, perhaps you could find, depending on the variation in the social factors, greater environmental variance would be my guess from what little we know in other, younger populations.

That said, this particular sample, most of these individuals had fairly low education, and it was only later, of course, in later cohorts, that you see greater access to educational attainment than in this cohort. So we do find association, even still, with this cohort, with educational level and cognitive performance, primarily. So there is still some variation here yet with respect to social factors.

Q: So there's [an apparent decrease] in genetic [variance decrease] after [age] 65. Have you considered epigenetic changes?

Reynolds: Well, yes, we should consider that as a potential explanation. But, no, we have not actually

measured epigenetic influences *per se*.

Q: But there are some suggestions, actually, from twin studies, that there are differences with age...

Reynolds: Right. So, there have been cross-sectional comparisons of twins, and you see greater divergence in methylation patterns with age with identical twins. So that definitely could be a factor. That may be explaining some of the changes in variation with age, as well.

Q: Why does it go down after '65, except for memory?

Reynolds: Right. That's interesting, and I've looked at the worldwide evidence from cross-sectional and longitudinal studies to see if you could also ascertain such a pattern. And it does look to be the case that you see some non-linear change in genetic variance across studies, such that you eventually see an increase in genetic variance after age...

Perhaps it's some particular time period in which you see genes such as ApoE4, which does have a window of particular effect, between about 65 and 75, or 60 and 70 -- perhaps this may be one of other factors that might be explaining it. We see it across all our memory measures, and the information measure, which does have some semantic memory aspects to it. So we think it's probably a true finding, but that's our best guess at the moment.

Q: I have a very general question to you, Dr. Hendrie. Is the increased biomedical risk in the African American population...does that account for their increased risk of Alzheimer's disease? Do you have a general sense of that?

Hendrie: I think that was our original hypothesis -- that, in fact, the difference between Yoruba and African Americans may be associated with the additional risk for cardiovascular disease that the African Americans have. Having looked at the data, now, I'm not sure that really does explain the whole differences. There have been comparisons in other studies between African Americans and Caucasian -- or European/American populations -- and while the result was this increased risk for hypertension/ diabetes, in most of the models that try to explain variance, it doesn't explain the difference in variance between the two populations.

So, I would be kind of cautious about that. I should emphasize that the Yoruba are not healthy, you know? The elderly Yoruba die at a higher rate...at an increased rate than elderly African Americans, but they die of different things. They die, still, of infectious diseases, and certain types of virulent cancer and so on.

So I think it is kind of complicated. And it's probably, as our last speaker said, involved in different models that occur as people get older, with genetic and environmental influences. You're right in the sense that the impact of disease is probably a great deal more as people get older than it is in younger populations, and so there's some difference in mix. But at the moment, we're not sure we can justify the original speculation about cardiovascular disease.

Q: Do you have a handle on psychosocial factors? You discussed social circumstances, and material circumstances, but I'm wondering ___ I would guess ___ might be at greater risk among African Americans...

Hendrie: Well, actually, Dr. Hall is particularly interested in psychosocial factors, and we have a huge component that she is just about to publish that will tell you all about psychosocial factors in both populations. And it does affect risk. You see it more, I think, in the African Americans, maybe because some of them have a bigger variance than they are...

So, for example, low education is a risk factor for African Americans. That doesn't seem to have much effect in Yoruba. Although it depends, a little bit, on how you look at the outcomes. So that the small proportion of *this* cohort of Yoruba, that had any education, were at equal risk for developing incident Alzheimer's disease.

There was an interesting interaction we found in the African-Americans -- again, Kathleen has published this -- between education, and where people lived up until the age of nineteen. And the interactions said that the education effect was particularly involved with African Americans who lived the first years of their life in Southern rural communities, and then came with that huge African American migration to the big cities of the North -- Peggy will tell us more about that -- in the '40s and '50s.

So the education influence was coming from this population who spent their childhood in these

relatively poor -- very poor -- deprived areas of the rural South…which makes us think, for instance, that the education component is really a marker for some sort of deleterious childhood experiences.

[*Applause.*]

Lindau: Thank you so much for moderating. Unfortunately, as you know, Richard Suzman is not able to make it. We have a little bit of buffer time. I don't think Lis and I will need fifteen minutes to close, so we'll still end on time.

I want to thank Lindsay Chase-Lansdale for coming from Northwestern. She will moderate the next session. I've also asked her, since it's a wonderful opportunity that she's here, to briefly introduce the group to the Center for Cells to Society at Northwestern, so you're familiar with the activities there, some of which are coming up next week. So, Lindsay, I'll turn it over to you, and we're looking forward to this next exciting session.

Chase-Lansdale: Thank you, Stacy. It's a great opportunity to have a chance to briefly tell you about C2S, or Cells To Society, as we call it for short, and then introduce our wonderful speakers.

Presentation: ***Future Directions in the Integration of Biological and Social Measures, from Theory to Analysis***

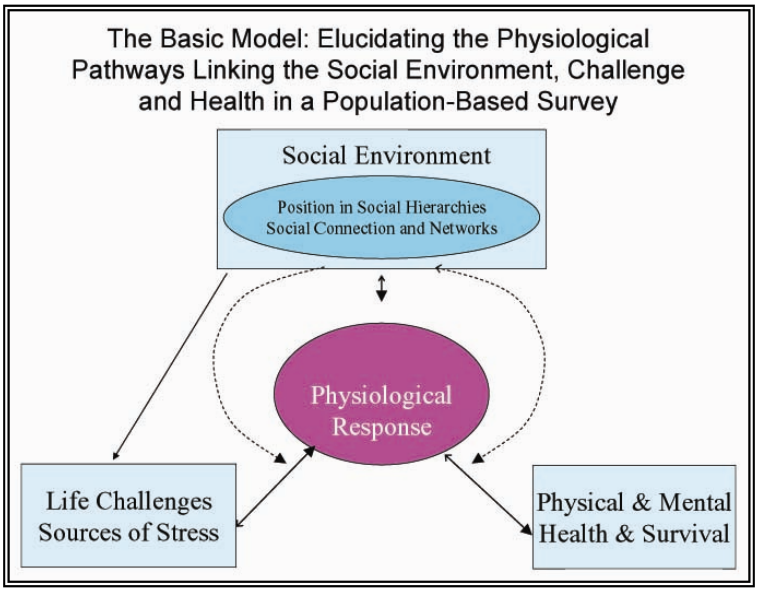
Noreen Goldman, D.Sc (Princeton)

Thank you very much for that introduction, and thank you for having invited me to this workshop. What I'm going to do here is talk about a project I've been involved in for the past decade, and to do this from the perspective of a social scientist. I'm a demographer who has moved in the direction of incorporating biomarkers in my research. I will try to raise some questions about what have we learned by doing this, and the kinds of problems that we faced, and I am sure, the problems that we are likely to be faced by many such enterprises.

My project was carried out in Taiwan. It looks at SES, stress, health, and the physiological linkages amongst these factors. And I list here a group of collaborators, notably my collaborators in Taiwan; the NIA, who has funded this project; and my co-PI, Maxine Weinstein.

So I'm going to try to do three things: First, give an overview of the project and the fieldwork. The project is called SEBAS, for the Social Environment and Biomarkers of Aging Study, and it's also because I love Chilean sea bass, and I knew I could at least remember the acronym if I couldn't remember what it stood for. The second is to give a brief summary of recent findings pertaining to the physiological linkages among social factors, stress, and health. And my third goal here today is to talk about the problems that come up in doing this kind of research.

So here's the basic model -- a very simplified form with only a few arrows -- of a kind of diagram many of you have faced in your own research. We know that the social environment -- SES, social ties -- affect health, where health is survival and physical and mental wellbeing. We know that stress -- and I'll define stress a little bit better later on -- but let me just say now, life challenges affect health.



We have some information of the physiology from clinical studies, but what we're really missing is a kind of data enterprise that puts it all together, that collects extensive information on sources of stress, social factors, health, and the potential physiological response to social factors and stress that mediate the relationship with health. So that was our goal in the late 1990s, when we undertook the survey in Taiwan.

Our survey built upon surveys that had already been in the field, carried out by what is now the Bureau of Health Promotion in Taiwan. They began in 1989. They were actually the Taiwan Provincial Institute of Family Planning, but once almost everybody was using family planning in Taiwan, they knew if they didn't move directions, they were out of business. So in the late 1980s, they recognized the growing elderly population, and moved into the area of surveys of the older population.

And every three years, roughly, they interviewed a population starting with 60-and-over, and they had refresher cohorts. And then in 2000, we worked collaboratively with them to take a sub-sample of this population, and have an extensive biomarker collection that I'll very briefly describe in a minute, starting with people 54 and older, with face-to-face interviews and biomedical data collection. And these people would continue to be followed up as part of their enterprise. We just came out of the field a couple of months ago, having done a second round -- what we called SEBAS-2 -- following the people from the first round and, again, having a refresher cohort.

So that's a brief summary of the fieldwork. What we did in 2000 was that we interviewed people in their homes -- actually, local public health nurses did that. Then, several weeks after the home interview, people were brought into the hospital for an exam. That was the preferred way of doing venous blood collection in Taiwan. The night before the hospital visit, they took

an overnight urine sample.

The respondents were then picked up by a staff person, taken to the hospital for a full health exam -- more extensive than what they would have been

Summary of Survey Characteristics
Main Interviews: Surveys of Health & Living Status of the Near Elderly & Elderly

Year	Target Pop	Type	# interviews
1989	60+	Face	4049
1993	64+	Face	3155
1996	67+ 50-66	Face	2669 2462
1999	53+	Face	4440
2000 **SEBAS I**	**54+**	**Face** **Biomedical**	**1497** **1023**
2003/04	57+ 50-56	Face	3778 1599
2006/07 **SEBAS II**	**53+**	**Face** **Biomarkers**	**1284** **1036**

entitled to through the national health insurance exam -- which included blood collection and measurements, and abdominal ultrasounds, and physician check-ups. It was national, and part of the challenge was to be able to do this, not on a small scale, but nationwide, in a very hot climate.

So, there was essentially one hospital for each of the primary sampling units in the survey. Ninety-something percent responded to the interview, about seventy percent to the hospital part, and we've analyzed extensively the non-response rate. Interestingly enough, it's not surprising that you lose the most disabled from this enterprise. In fact, we have a series of questions whereby we don't let people come to the hospital if they have any of a number of conditions.

But you also lose the healthiest, because these are people who feel there's no benefit, they don't need a health exam -- they're healthy -- or their work schedule is so tough that the idea of taking off a half-day to do this enterprise is impossible. So, there's a little bit of each tail that gets lost in the enterprise, but, on average, the health status of those we missed is similar to those that we have.

In the home interview, we have questions that weren't asked before in these earlier waves -- for example, questions having to do with perceptions of stress, and stressful experience. An earthquake came just before we got into the field -- a major one that delayed our fieldwork -- so we were able to look at some of the consequences of the earthquake.

Here we have a list of some of the biomarkers we collect: From the overnight urine, we have cortisol, epinephrine, norepinephrine, dopamine. From the hospital visit, we have fasting blood, measures of cholesterol, glycosylated hemoglobin, IL-6, IGF-1, DHEAS. As I mentioned, we also have physical exams and lab items that come out of standard physical exams. We collected APOe genotype, not because we had any plans to use it, but because the review panel suggested that we add that to the enterprise. And we had very high compliance on this.

I have a little picture here of cortisol concentration by the time of day. Most of you are familiar with this. But it's a reason why one wants to get some kind of integrated value, and so we have 12-hour overnight measures of urinary cortisol.

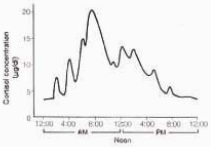
Just a few pictures, because pictures are nice to look at, but they also reveal the kinds of complications that arise with biomarker collection, which was a new type of data collection for all of us, and life got much more complicated. Our investigators spent months designing ways of collecting 12-hour urine.

Hospital Visit/Biomarker Collection

Similar to MacArthur Study

- **Overnight (12 hr) urine collection**
  - Neuroendocrine markers (cortisol, epinephrine, norepinephrine, dopamine)
  - 12 hr to obtain integrated resting values
- **Hospital visit**
  - Measurements (height, weight, waist, hip, blood pressure)
  - (Fasting) blood test items (DHEAS, cholesterol, glycosylated hemoglobin & fasting glucose, IL-6, IGF-1)
  - Physical exam & lab test items similar to National Health Exam + abdominal ultrasound + APOE genotype
  - Very high compliance
  - Respondents received reports on clinical markers & doctor's exam

Cortisol concentration by time of day



And somebody else brought up the dry ice issue. That was a big one for us, too -- where we would get it, where we would store it. Most people don't have large refrigerators. If they do have refrigerators, they're not big, and they're not about to store overnight urine in them. Again, this is a very hot climate, so a lot of work

went into these devices to collect and store the urine; it was not a simple enterprise.

And here is an example of one unintended stressor that hit... This is a middle school, not that far from where the earthquake hit. That *was* a level running field. So, as much as you plan, not everything is in one's control.

"...the biomarkers do matter; in fact, they matter a lot.."

We went back into the field in 2006-7. We followed up with people. We had our refresher cohort. We added several measurements we hadn't the first time around. We took blood pressure at two points: We had taken it in the hospital -- I think we had white coat hypertension -- so we had another regime this time that took blood pressure readings in the home, along with lung function and grip strength. We added some biomarkers -- C-reactive protein that many people have talked about in their own enterprises, fibrinogen, another genetic marker -- 5HTT related to depression. Telomere length is also something that we are measuring now.

Interestingly enough, we tried two things that didn't work in our pilot for the second round. We tried to take a 20-minute EKG -- something that would be less trouble than a 24-hour Holter monitor. That didn't go over big. Respondents had a lot of trouble with adding one more regime, sitting still for that long of a time, being very uncomfortable with the fact that that this wasn't the kind of EKG they had experienced when they had check-ups. Similar problems occurred with salivary cortisol.

Taiwan, interestingly enough, is a highly educated population today, but among the older cohorts, almost half of the women are illiterate. So that meant when we left instructions, for example, for the 12-hour urine, or the salivary cortisol -- we relied primarily on pictures for many respondents. And trying to do those things at the same time -- having people collect overnight urine; salivary cortisol in the evening, again in the morning; having them be ready to be picked-up to go into the hospital in the morning for an exam -- that was not a comfortable load on respondents.

We also introduced some new questions on sleep, on trauma, on caregiver stress, pain, relaxation. We're out of the field now, but the bigger enterprise will continue to be carried out every few years by the Bureau of Health Promotion, and there are linkages to survival data through the national registration system.

This is a picture of the crew in one of our hospitals for the second round of fieldwork. What's always impressed me, and made me realize I was so happy to be working in Taiwan and not the U.S., is that we would arrive, and hospital administrators would come down; they would say, "Before you do anything else, come have coffee." A group of physicians would greet us. The kind of hospitality and support in this country was absolutely overwhelming for the entire project. And we would get ample space, such as a portion of an emergency wing, to operate our survey out of, for the few days we were in every primary sampling unit. You can even see a welcome sign of our project that was put up in each hospital. I just can't imagine that happening in the United States.

Here's an example of a home we visited.

So, what I'm going to do in the second part is try to give you a very brief summary answer to three of the major questions that we've been asking in the last couple of years while analyzing these data...And they all relate to these biomarkers...I'm not going to take you through the actual statistical models -- there's not enough time to do that -- but let me just say a couple of things.

When we analyze biomarkers, we've done it in many different ways. We either looked at all the individual biomarkers and then defined what we mean by high-risk -- whether it's a quadratic function, or a low- or a high-tail. We've also done the analysis by using summary scores. Allostatic load is the conventional term for putting together biomarkers into a kind of count index. We've looked at many different kinds of modifications of the conventional index. We've moved away from calling it "allostatic load," to physiological dysregulation. But these different analyses have formulated the biomarkers in many ways so that we can feel confident about the association of biomarkers with these other social factors.

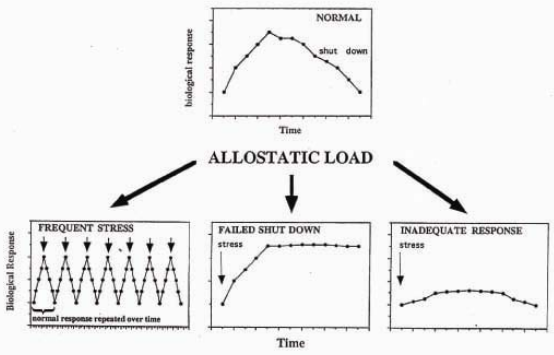
The first question that we asked is, "Are these biomarkers -- these biomarkers are ones that have been posited to be part of the response to chronic stressors -- are they associated with health and

mortality?" Secondly, are they associated with measures of stress in the environment? And third, are the biomarkers associated with socio-economic status, and if so, do we have support for the argument that "stress" accounts for the kinds of social disparities that we see in health outcomes? And that's a tall load for about ten minutes, even for a New Yorker, to speak at that rate.

Most of you are familiar, I'm sure, with the notion of allostatic load, but for the few who aren't, let me just say in a couple of sentences that the idea behind this is looking at the effects of chronic stressors on physiological function and, ultimately, health. If one is subject to an acute stressor, one's biological response, whether it's cortisol or blood pressure, goes up, and when the stressor is gone, comes down. And all is fine. And that's a very adaptive response.

The issue is, when the stress is chronic -- chronic meaning it's there for a long period, or it's constantly coming all the time -- that's where the notion of allostasis, the body being able to respond and adapt to that kind of change, falls apart; and it's what McEwen and Stellar have called "allostatic load" that comes into play. There's a kind of physiological dysregulation that occurs.

Allostatic load: Repeated or chronic environmental challenges result in dysregulation across various physiological systems (***neuroendocrine***, ***immunological***, and ***sympathetic nervous system*** markers become out-of-normal range), which can lead to chronic disease.



Biomarkers that are meant to adapt, don't adapt. They either continue to stay high -- your blood pressure has gone up so often, it no longer comes down -- or you no longer are able to mount an appropriate of stress response. Your cortisol doesn't respond adequately in a stressful situation in which it's meant to. And so what you have -- this kind of wear and tear on the physiological systems -- leads to biomarkers being out of what are called "normal operating ranges." And it's that kind of biomarker dysregulation that we measure in our enterprise.

So, question one: How well do the biomarkers predict mortality, compared, for example, to the other kinds of predictions demographers are used to using -- people's self-reports of health... And I don't mean just a simple question, would you rate your health "Excellent, Very Good, Good,…" but self-reports in terms of disabilities, presence of chronic disease, pain, smoking, all sorts of questions like that.

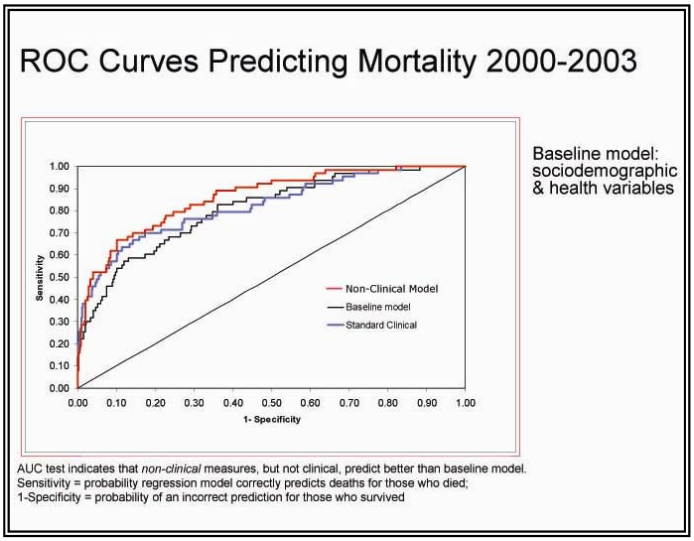
And to the extent that biomarkers matter, is it the clinical factors -- is it blood pressure and cholesterol, the kinds of parameters a doctor will measure -- that matter the most, vis-à-vis the kinds of markers that we've collected that are not clinically measured, like cortisol or epinephrine, or DHEAS?

So the results that I'm going to show you are based on using 2000 data -- we've not yet had a chance to look at the 2006-7 data; they just came back -- and following survival for three years after that. We just received three more years of survival data, but those results are still very preliminary.

The answer is that the biomarkers do matter; in fact, they matter a lot. Even in the presence of extensive controls for self-reports of physical, mental, and cognitive health, the biomarkers predict survival. They do so astoundingly better than self-reports of health. And the non-clinical factors -- the neuroendocrine and the immune markers -- seem to matter at least as much, if not more, than the clinical markers, such as markers of cardiovascular disease, diabetes, or metabolic syndrome more generally.

Just to throw in a little data, because it's so hard for me to give a talk without a graph or a number, having been raised partially as a statistician... Here is a graph that shows you...using regression models to predict mortality for this three-year period based on three statistical models: A baseline model that just looks at sociodemographic factors and people's self-reports of health; a second model in blue that adds the clinical markers of metabolic syndrome to that; and a model in red that doesn't add those standard clinical markers, but instead adds the non-clinical markers. That is, the neuroendocrine immune markers.

And you can see that the higher the graph -- the greater the area under the graph of the ROC curve -- the better the prediction. And in fact, the area under the red curve is statistically bigger than the area under the blue, which was a little bit of a surprise to us. And



you can see the improvement in prediction by adding biomarkers to a much more standard social science demographic prediction model.

So the second question that I ask is, are biomarkers associated with stress? And before I use that word any more, let me define it. So, people refer to stress in three different ways in this figure. These three boxes denote three different aspects of what people mean when they refer to stress. One is chronic stressors in the environment -- kinds of life events, or traumatic events, that we think cause people to have this stress response.

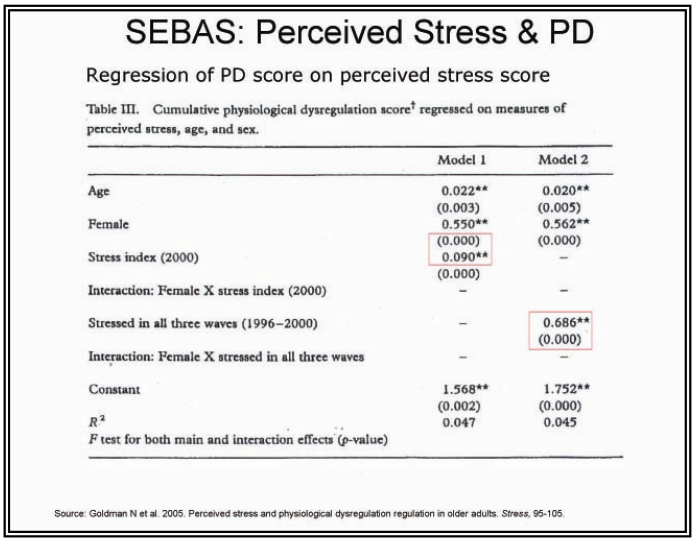
The second has to do with people's perceptions of stress. Scientists argue that it's not just the events that occur, but it's how people perceive them, and in the jump from actual stressors to perceived stress, many factors -- such as personality and social ties, access to resources -- come into play, because, for example, people who are more optimistic, have better locus of control, have more resources, and more social ties -- are likely to perceive stress less severely in the face of a given type of chronic stressor. And then there's the third interpretation: stress in terms of physiological response to stressors -- what we call physiological dysregulation.

So we'd like to separate these different notions of stress here, and look both at chronically stressful events -- what I'll call stressors in terms of their impact on physiological dysregulation -- and look at perceptions about stress and the impact of those.

So, a brief summary of findings... First, with regard to perceptions of stress. In this part of the analysis, we looked at 16 of the biomarkers that I showed you a few minutes ago. Five of them have significant associations with an index of perceived stress that we constructed based on responses in 2000.

And so you may say, "That's good." I don't know. We figured that was probably not so great. All of these markers are posited to be part of the stress response. I've listed here the markers that were significant -- cortisol being one of them -- but many biomarkers revealed no significant association with perceptions of stress -- blood pressure, markers of obesity, total cholesterol.

When we looked at this, not in terms of individual biomarkers but a dysregulation score, it was significantly associated with perceptions of stress. But I think if you focus on the significance of the results -- and many people do that; -- you'll walk away saying, "Yes, we found what we expected." But, if you translate the results into magnitudes of effect, you find, for example, that people who perceived stress in three consecutive waves of the interview had about two-thirds of a point higher on the physiological dysregulation score than people who did not. That's



two-thirds of one more biomarker out of range. So, the dysregulation score is a count of how many biomarkers were out of normal operating ranges. I think that this estimate is modest in terms of its impact.

And at the bottom of the slide are the kinds of question that were used to ascertain perceptions of stress. Respondents were asked about each one of these potential stressors. Do your finances make you feel distressed or anxious? Do your social relationships...?

I guess I threw in one more set of numbers. These are the actual regression coefficients, with the boxes a little bit off. But you can see that the stress index is significantly related to physiological dysregulation in Model One. And in Model Two, instead, the index was replaced by an indicator for people who were stressed in all three waves; that too showed a significant association. In fact, in models with both measures, both show a significant association with physiological dysregulation -- significant but modest.

If we now look, not at perceptions of stress, but at actual stressful experiences, I will say that it's one of the hardest analyses I have ever, ever done. The difficulty arises because in every single wave, people are asked about many potential stressors. There are also potential moderators of stress asked in every wave -- SES, social support, locus of control, optimism, on and on and on.

And trying to think about how you put that all together in a statistical model and still have reasonable statistical power, poses a very difficult problem, particularly, since our theoretical model entails many potential interaction terms between the stressors and the moderating factors. The idea of interactions has come up repeatedly here because of the potential importance of gene-environmental interactions, but it's just as important in the present context.

Although we view these social and personality variables as potential moderators of stressful experience, many researchers will just put in income or education as direct effects in a statistical model. We believe this is not a good reflection of the actual process because, given a particular set of traumatic or stressful experiences, they may act to create higher or lower levels of perceptions of stress and thus different degrees of physiological impact. And it's very hard to model this process statistically, in terms of complexity and statistical power.

So, we've done it every-which-way. I think the paper is finally about to come out. But it took a long time to get off the desk, until we convinced ourselves that that answer was pretty robust. And the answer is that the effect is really small. If you're trying to look at different domains of stressors, there's nothing significant going on in terms of association with physiological dysregulation. If you create a kind of score of the number of stressors experienced in the environment, it is statistically significant for the people that we would call highly vulnerable. And those are people low on social position, social support, and internal resources -- a compendium of the kinds of moderating factors I've talked about. And the effect is pretty modest again.

"Although we view these social and personality variables as potential moderators of stressful experience, many researchers will just put in income or education as direct effects in a statistical model. We believe this is not a good reflection of the actual process because, given a particular set of traumatic or stressful experiences, they may act to create higher or lower levels of perceptions of stress and thus different degrees of physiological impact."

We examined eleven potential stressors, including severe responses to the earthquake, in terms of damage, severe financial difficulties, loss of spouse, loss of kid -- the magnitude of the coefficient indicates that one additional stressor out of eleven was associated with an increase in the physiological dysregulation score of about 0.1. In other words, it would take ten stressors to generate one more biomarker out of range. So, again, it was statistically significant, but for the highly vulnerable, not for others, which supports the idea of these moderating factors.

Finally, are the biomarkers associated with SES? And the answer here is that a large number of the associations between what we call "at-risk biomarker values" -- high blood pressure, high cholesterol -- and income or education, or other SES measures are insignificant. The patterns are not consistent, and in particular, lower income or lower education, were not significantly associated with higher cortisol. So, not surprisingly, if you go the next step and you ask, "How much do the biomarkers account for the association between SES and health?" -- measuring SES in various ways -- the answer is, relatively little because you don't see a consistent association between at-risk biomarker values and SES.

I've spoken with my colleague in Costa Rica who's done a similar enterprise, who finds exactly the same thing. And I'll get back to this point about Costa Rica and Taiwan in a minute, but the main point is that there is no consistent association between SES and biomarker values. And to the extent you see anything, the cardiovascular risk factors actually show more consistent patterns with SES than do neuroendocrine or immune markers. So, at least from these data, there's little evidence that chronic stress via sustained activation of the HPA, or the sympathetic nervous systems, is an important mediator in the association between SES and health. And I'll give you the caveats that obviously come with this in a minute.

And here's just an example of some numbers. This comes from a paper recently published in Population Development Review. And in the box, I've outlined the coefficients that come from modeling education and occupational status of father. In Model 1, that only looks at social and demographic variables, and Model 2, that adds physical and mental health controls to that, you can see that the absolute value of the education coefficients go down in magnitude. That is, you account for some of the education differences in Model 1 by adding self-reports of health to Model 2. And then, when you add biomarkers to Model 3, the coefficients go down a little bit, but not a whole lot. I mean, the inclusion of biomarkers does not really account for very much of the differential.

SEBAS: Education & Mortality Effects of Adding Biomarkers

Logistic models of 3-year prob of dying

Model 1: Social or demographic variables

Model 2: Self reports of physical and mental health added to (1)

Model 3: Biomarkers (linear and quadratic) added to (2)

Variable	Model 1	Model 2	Model 3
Demographic characteristics			
Age	0.1020**	0.0952**	0.0807**
	[0.0202]	[0.0227]	[0.0279]
Male	1.0522**	0.9651*	1.4901**
	[0.3493]	[0.3968]	[0.4915]
Had spouse/partner in 2000	−0.4337	−0.3464	−0.2641
	[0.3074]	[0.3166]	[0.3552]
Mainlander	−0.2808	−0.2523	−0.1795
	[0.3789]	[0.3890]	[0.4389]
Social and economic variables			
1–6 years of education or literate	−0.5034	−0.3947	−0.2935
	[0.3368]	[0.3502]	[0.3908]
7+ years of education	−0.7560+	−0.3363	−0.2330
	[0.4447]	[0.4778]	[0.5498]
Paternal occupational status index	−0.1241	−0.0962	−0.0876
	[0.1104]	[0.1157]	[0.1144]
No. of social ties with nonrelatives	−0.0408	−0.0324	−0.0401
	[0.0269]	[0.0256]	[0.0311]
No. of social activities in 2000	−0.0156	0.0258	−0.0316
	[0.1557]	[0.1597]	[0.1761]

Source: Turra CM et al., 2005. Determinants of Mortality at Older Ages: The Role of Biological Markers of Chronic Disease. *Pop Dev Review*, 675-698.

So, in summary, are the biomarkers associated with health and mortality? No question. Strongly. Are biomarkers associated with measures of stress -- both perceived stress and environmental stressors? I'd say significantly but modestly. Third, are the biomarkers associated with SES? A few, but the direction is not consistent, and there's not much evidence that the biomarkers mediate the relationship between SES and health.

According to your time, I've got five minutes, and I'll spend the last five minutes saying what I think is going on here. Why do some results match expectations, but some do not? And I think this leads to five important questions that we need to think about if we are to use these kinds of large-scale social science surveys with biomarkers added to them, to study the linkages between stress and health.

So the first question is how well can we, in general, measure chronic stressful experiences in a standard type of questionnaire? What we've done -- we've done a little more in the second round of our data enterprise than the first -- is we asked people about traumatic events, life events, repeated daily hassles. However, this is a survey of people in the 50s and older, so to some extent you're asking people to recollect the past, and how well can people report this? For traumatic experiences, it's probably not that hard, at least to recall the occurrence. For some other forms of stressful events, I think it probably is. How much can you ask people about intensity, duration, and timing of these stressful events, when you're asking people to report this retrospectively?

Some of the same issues pertain to perceptions of stress. We're actually evaluating the Cohen Perceived Stress Questionnaire in this round, but we've incorporated many other ways of trying to get at that kind of information.

Second -- equally big question -- is how good are our measures of biomarkers? And undoubtedly the answer is, not great. We're starting this kind of enterprise. Have we gotten the important biomarkers? Well, we're all getting biomarkers that are relatively easy to measure, and relatively non-invasive, what you can collect in this kind of one-or-two-or-three-times shot.

Do we have the correct measure? We're typically going for averages. We tried variability in terms of salivary cortisol. We tried it in terms of heart rate. It was much harder to incorporate into our design, and that's

probably a problem. So now we have two measures of biomarkers; we started with one. In some cases like blood pressure, we do have a series of readings, but that's still a series of readings at essentially one point in time, that time being a half-hour. That's probably vastly inadequate, and it's probably unrealistic to think that a single time measure will give us a very good representation, for example, of the physiological response to chronic stressors.

There's a big debate in the literature about using these kinds of scores of allostatic load. We personally are pretty unhappy with the idea of a score, whether it's a count, or some much more sophisticated way of looking at this that we've tried to carry out. And part of the problem is, I think, that often important associations can get obscured. And part of the problem is that these scores are not very theoretically driven. I could easily go on for the thirty minutes saying nothing but problems associated with summary measures of allostatic load, so I'll stop right there.

And the last bit under point two is how much respondent burden can we impose in this kind of setting? For example, I was in Moscow a month ago, working with a new survey that's taking place there that's using the Holter monitor. That's actually the main piece of their enterprise, and a very exciting enterprise. I know that wouldn't have worked in our case. And, I think one has to adapt to different samples and different cultures, and think about how much respondents can handle. In their case, it was a 24-hour monitor that has been a major part of the fieldwork and the cost.

So, number three -- How well do we measure variables -- variables like personality, characteristics, coping skills, locus of control -- that moderate stress? These were also very hard to measure in the Taiwanese context. We tried a module on optimism, and gave it up retaining only a single question, because each question was 100% correlated with every other one. People just were uncomfortable, and if they answer on a scale of one to five, they gave four to every single answer. Or if was reverse-coded, they reverse-coded it, mentally. So it's something we just couldn't tap, even with modifications to Western-type questionnaires.

And apart from getting the right questions, do we have the statistical power to look at what are essentially interaction terms? And do we have statistical models suited to this complexity? You can't just take a load of interaction terms, throw them into a statistical model, and see what comes out. That's not going to have very good predictive power in another sample other than your own sample. And I think that the conventional regression models are not so well suited in dealing with these kinds of problems.

Fourth out of my five -- Are we looking at too old ages? So we're looking, in our case, at 54 and older. I was very sensitive to this in the Moscow survey, because Moscow is a country, as many of you know... Moscow isn't a country, but Russia is a country…

[*Laughter*.]

Moscow is actually better in this regard with lower mortality than Russia as a whole. Russia is a county with very high mortality; one of the few non-AIDS-driven countries that had major reversals in survival in the last couple of decades. And so, by the time you got to the age of this survey, many, many people had died. And they did so differentially, by SES. Russian demographers are very interested in social disparities in health, but you lose the vast majority of the people of lower SES if you restrict the survey to people say 60 and older. As many of us know, regardless of country, social disparities in health tend to weaken with age. And many of these surveys come in at a point when this association has already weakened. So maybe this is why we're not seeing as much as we expect.

And a fifth, and a critical question -- one I've asked myself a lot -- is... So when we first got some of these results in Taiwan, they seemed more modest, for example, than we would see in England with Whitehall, or that we would see in some of the U.S. studies, such as MacArthur. The more I looked at some of the U.S. studies; however, I saw they weren't as radically different as we thought. But I think it's fair to ask the question about whether we expect the same result across countries.

So part of our thought is that there is less variability in SES in Taiwan and Costa Rica, the other country that had the same results that we did. Fewer older people are socially isolated. There are very strong family ties. There's a lot of living together in households; and even though multi-generational households are becoming less prevalent, there are still very strong connections with kin, and with neighbors and with friends. There is a lower incidence of many traumatic events -- violent crime is very rare in Taiwan.

And so you can ask whether the associations between measures of stress and biomarkers are stronger in some societies because the frequency of these extreme forms of stressors are more common, and while the pattern may be the same here, we don't have the prevalence of extreme stress. But it becomes very difficult to replicate and evaluate the magnitude of the result and assess whether or not the findings are consistent across multiple places. We don't even know how much we would expect the relationships to vary across cultural contexts, and how much we would expect to be able to confirm our findings by looking cross-culturally.

So the bottom line... Everyone asks me this question, and I figured I might as well just put it out there, and struggle with it in front of you, rather than have somebody ask it. Does the inclusion of biomarkers in household surveys help us to understand the mechanisms underlying social disparities in health, and the role of stressful experience in physiological dysregulation in health?

And I would say, we don't have a clear answer. I think we're still at the start of this enterprise, "we" being social scientists. Clinical scientists and epidemiologists are some steps ahead of us. I think we're moving in the right direction. I think we'll improve our data collection and our analytic strategy as time goes on, and get a better answer to that question.

[*Applause.*]

Q & A with Dr. Goldman

Q: Hi, Noreen. I just had a quick question. I wasn't sure about your last comment about the… Should we expect large variation across settings? And you found in Taiwan and Costa Rica that, really, there's less variability in SES. And I didn't catch if you gave an explanation of why you think there was less variability in SES in Taiwan, for example. Due to attrition or to...?

Goldman: No, I think you just see less extreme forms of poverty in these societies. You have much greater government role... The kinds of extreme wealth are just coming to Taiwan. You simply don't have that extreme rich class, and that very high proportion of living in poverty in a much more socialized system that...

Q: In Taiwan and in Costa Rica...?

Goldman: And in Costa Rica; Costa Rica probably even more so. You have much more access, for example, to health care. Both countries are on national health insurance systems. So it's not just ___. I think they're very different societies, where you don't have that "underclass" to the extent that you would see it in the U.S.

Q: I want to put a hypothetical to you, Noreen, that might be useful. And that is, have you experimented with the following question? How poor would the measurements, say, of stress or of physiological dysregulation have to be for you to be able to maintain the hypothesis that they mediate the effects of SES on health? It seems to me that that's the kind of counter-factual that might be really helpful to you, even in the absence of models that would directly estimate the quality of the measurement of those intermediate variables.

Goldman: That's a very interesting question, and I've thought about that. The problem is, I don't know where to begin thinking about it. I know the biomarkers are measured poorly. You talk to a biologist and they'll say, "You did this on one day? Are you totally nuts?"

And so if I had to place a guess, I'd say that the biological measurement is weak in terms of thinking we can come in there at a given point in time, and capture something like a chronic experience. You know, it's much cleaner to think about response to acute stressors. So you give people public speaking tasks, and you have them subtract sevens at the same time, and you throw a hostile audience at them, and you measure their cortisol going up and down.

And it's much harder to do this... So you have people taking medications that are masking some of these responses. And so you have measurement error in at least two very, very critical places. I'm just getting a little piece of the biomarker story, even if it were an accurate reflection of it.

Q: But suppose you pulled something...just yanked a common factor out of the repeated measurements of

the biomarkers? Probably an incorrect way to do it -- that assumes no true change -- but that would give you a kind of an extreme assumption under which to look at other...behavior of other aspects of the model.

Goldman: That's an interesting question. I will think about that. I'll talk to you more about it, too. Yes, that's an interesting question.

Q: A little bit more about alternatives to simply the count combination of your biomarkers... It kind of relates to the last question. What is the optimal way to integrate these series of individual, somewhat unreliable measures?

Goldman: I wish you had been in Boulder during the last few days. I gave a two-hour lecture on exactly that question. There was a course on biomarkers, and one lecture pertained to how different people measure allostatic load. I can actually give you the PowerPoint for that if you like.

To say it in a sentence or two, there are many different kinds of scores that people have used; the conventional score just looks at high- and low-tail values. You could say, "Let's look at both tails. Let's not weight things equally, and just add them up." Let's think about weights. Canonical correlation procedures use survival, or some health outcome, and to determine how to weight the different markers."

There's a procedure called "Recursive Partitioning" that says, "Throw the whole idea of a score out the window, and let's think about identifying biomarkers that optimally distinguish those who survive and those who don't survive." And let the procedure also tell you where to cut those biomarkers -- don't impose an arbitrary highest-risk quartile, or highest decile, but estimate what you think the cut is for high-risk. And that's about as much as one long run-on sentence can do.

Q: I guess that one of the things that's interesting is you're saying that... I mean, your conclusion is you have the measurements but now, looking back on it, you're not sure whether you have the degree of S... So you have an answer to a question, which is, it doesn't look like the measurements that you did explain the SES associations with health. But now what you're saying is, well, maybe you don't have the SES associations correctly; maybe we didn't measure these things correctly, so the...

Goldman: I think we have the SES measurement correctly. The question is whether what we've captured is chronic stress.

Q: Well, it's not "correctly," but you said you weren't sure that you had enough of a disparity in the society in SES to measure this. And it seems to me that these are questions that people really have to think about before they go out and do the measurements. Because when you get to the end and you don't find an association...

I mean, I think people have to consider… A lot of the things that we're looking at are things that everyone believes very strongly *should* happen. And when they don't happen, I think the question is, "Do we have enough information in the studies to conclude that a negative result is a true negative result?"

And I think that that's a really important question when we're going to invest a lot of money in doing these biomarker assays, because it might be better to do one really well, rather than to do a whole battery of measurements where, at the end of the day, you're not sure that you've captured the true physiologic trait that you thought you were measuring. Or, if the degree of socioeconomic status within the project is not large enough, then we're never going to be able to really say that a negative result is a true negative result.

So just wanted to say that I think that this goes into the issue of which studies should we be measuring these markers in. And, not to say we shouldn't measure any markers in any studies, but, you know, that we need to choose them very carefully, so that when we get both positive and negative results, we're sure that what we get, we believe in.

Goldman: Let me just say that's a very good point. I think it's hard to do that *a priori* in the sense that we see SES disparities in health. They are present in our population. So there's no reason to think that one couldn't account for them in this pathway.

I just raised the issue -- maybe these stressors don't have effects until you get to the extremes. That, in itself, would be an important form of knowledge. But I don't think, one, you could know... If you saw no SES disparities in health, I would say, yes, I mean, it's silly to be going down this avenue. But despite not having the extremes of social disparities, you simply see what's

called the "social gradient." At every step of the way, there is, in fact, better health among those with higher income, more educated.

Q: A quick question and a comment. I believe I read in one of your papers, you found low cortisol was associated with stress. Did you ever find that opposite finding in this Taiwan cohort?

Goldman: In several of our cases, we found correlations with both high and low cortisol.

Q: Okay. So that's what I wanted to comment on. I think that's very significant. And I just want to underscore your point about good measurement of these fluctuating biological systems. These non-clinical measures are neuroendocrine hormonal dynamic systems. And in your picture of allostatic load, you pointed out that we can have an overresponse or an underresponse. So when we talk about dysregulation, it's bimodal -- it can go either way.

I personally have a large file drawer of cortisol studies that were not publishable because I have both hypocortisolemia and hypercortisolemia responders. And I think part of the strategy...a strength of having the large samples that you guys do... My problem is, I can't cut my sample up into three tertiles, usually, with confidence that I'm looking at a low dysregulated. But in these large studies, it's important to realize that you have two ends that can be dysregulated. Especially with cortisol, but some of the anabolic hormones, as well…

I guess my one suggestion would be... In my clinical studies, I look at suppression. And this is a very good test that is a little bit more of a yes/no -- are they dysregulated. Whereas cortisol, there are no real standardized norms, and each assay varies. Whereas with suppression test, you give a pill at home. It can be a low dose, you take cortisol the next day, and you get a sense of if there's a problem with the feedback loop.

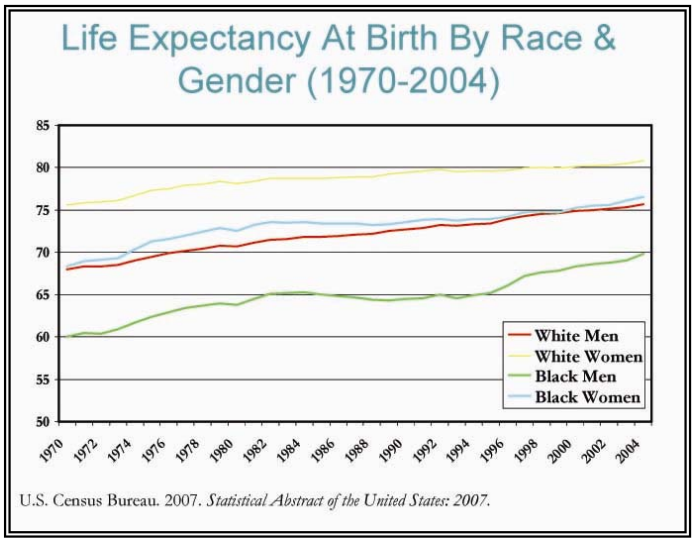
So I don't know why that's not used more these days. I know it's a little bit more intrusive than a diurnal monitoring. But anyway, I just want people to be thinking about the low cortisol end, too. And I know, for example, in Dr. Kurina slides, there might be...there was some, looked like very low cortisol. It might have been low-stress, but it might also, if you analyze it by stress, it might be that you see this other pattern of dysregulation.

Goldman: I'd like to respond also in the context of the earlier question on how you measure allostatic load, which I think some of those early measurements are really deficient in that there is recognition of one tail or the other tail but not both. Diastolic blood pressure is another very clear-cut case where so many of the problems erupt at the low end. For BMI among an older population, you're more likely to see problems at the low end than the high end. So I think we certainly need to adapt our measurement to recognition of this issue of two tails of risk for many biomarkers.

Q: Great! Thank you so much.

Arline Geronimus, Sc.D (University of Michigan)

Thanks very much. It's a pleasure to be here. I think this is a nice companion piece to Noreen's talk, both



because she saved me from describing certain things like allostatic load, but also because in this case, perhaps unlike her study population, we are talking about a population that experiences chronic stress with a big C. And I think that when we use words like stress -- whether it's acute, or chronic, life-event stress, environmental stress -- there are very big differences in life experience, when a person is being hit from every level, every day, in ways they perceive, and also, in ways they may have no clue are happening. I think, this is the situation for at least some groups of African-Americans in the U.S.

And so I think, another aspect we need to consider, when we think about stress measures, is that there may be variation in intensity and pervasiveness. And it is also important to think about *groups* of people and their structural locations in whatever society they're in, rather than view stress only from the vantage point of our being individuals where some of us have this many life-event stressors, and others of us that many. But how does being a member of a group that is constantly bombarded by stressors different from just being someone who, within your own group, might have more life-event stress than others?

Toward that end, I also want to put in a plug for mixed-method approaches that combine ethnographies, or qualitative interviews, with statistical work, to help understand what the experiences and social locations of the populations we're studying. So, with that backdrop, let me apologize, also. I've gotten the sense that for some of you, this talk is going to be too basic, and for others, maybe, who haven't heard of weathering before, I may not have planned to say enough. Hopefully, there is somebody here, to whom this presentation will be pitched correctly!

I think everybody here is aware that there are racial disparities in health in the U.S. You could pick almost any health indicator and see them. The example I'm showing is life expectancy at birth. And the racial disparities are fairly large and they are persistent. Unfortunately, that's the story, even though there have been mobilizations to reduce...or, I guess, we've now been implored to *eliminate*, these disparities, which we would like to do. But in fact, they persist, and there is even some evidence that they've grown over the period that we in the public health community have been charged with eliminating them.

I want to highlight the variation within race, as well as the racial health inequality. And, again, I could have picked a lot of different examples, but what I picked to give you some sense of this is the probability of survival to ages 45, 65 and 85 among people who have lived to age 16 to begin with, so these probabilities are not informed by differences in infant mortality at all.

You can see big differences between whites and blacks in the nation as a whole, but also between blacks nationwide and black residents of high-poverty areas.

Probability of Survival to Ages 45, 65, & 85 Conditional on Survival to Age 16 , Women in Selected Populations, 2000*

Survival Watts to Age	U.S. Whites	U.S. Blacks	Southside Chicago ‡	Eastside Detroit ‡	L.A. ‡
.45	.98	.95	.94	.97	.95
.65	.88	.79	.67	.70	.73
.85	.42	.32	.26	.27	.27

In this case, I've chosen Chicago, Detroit and Watts, though we've done the same analysis, now, for 1980, 1990, and these are 2000, for about 23 local areas and get, with some variation, similar findings.

And what I'd like you to notice is that, for instance, in Chicago and Detroit, black men are as likely to die by

age 45 as whites in the nation are to die by age 65. And that overall, 40-50% of black men in these areas do not survive to age 65. Only a very small proportion survive to age 85.

Causes of Excess Deaths, 16-65 year olds, Black Men, 2000

	S. Chicago	E. Detroit	Watts
EDR*	935	665	708
% Circulatory	**34**	**33**	**32**
% Cancer	14	12	14
% HIV/AIDS	15	16	28
% Homicide	**7**	**8**	**5**

•Per 100,000 relative to whites, nationwide and age standardized to the white age distribution.
Source: Geronimus, Bound and Colen 2007.

Besides the questions these statistics raise scientifically or in terms of social justice concerns, on a methodological level, I want to echo comments such as Noreen's that they point to: when you start a study or a survey with 50 or 60 year olds, you've already missed a lot of black men. Those of you collecting data might want to think about that. For black women, the probabilities of survival are higher at every age than for black men, but the patterns are similar, with black women in the high poverty urban areas much less likely to survive than for the nation as a whole.

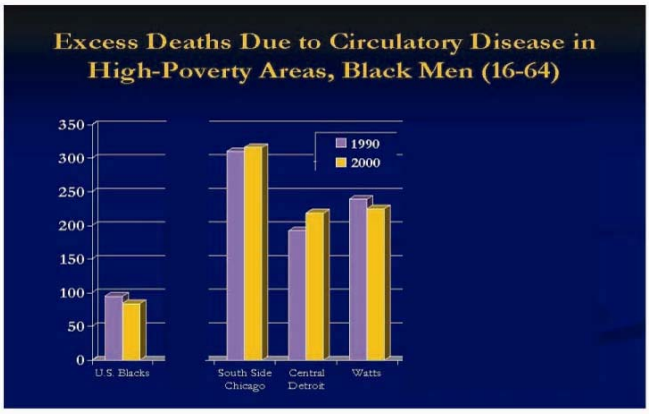
Regarding causes of these excess deaths, chronic diseases, circulatory diseases, and cancers are a big part of these. And I thought I would highlight, here, the percentage of excess deaths attributed to circulatory disease compared to homicide -- this is for the men -- because, certainly, tabloid images would have us believe homicide explains these high mortality rates. But, in fact, homicide is a much smaller proportion of these excess deaths than diseases of the circulatory system. And that's certainly also true of the women. More troubling still, there's some evidence that the contribution of chronic disease to excess death may be growing.

Just to give you some sense, this is a comparison of our findings for 1990 and 2000 in the same areas. In Chicago and Detroit, for men, the rates of circulatory disease deaths -- or excess deaths -- increased a little from 1990 to 2000, even though in the nation as a whole, they went down. In Watts, they also went down for men. Though for women in Watts, they went up.

And another example I want to point out is that for women in our Detroit area, and in some other areas we studied, we found that excess deaths went up between 1990 and 2000. And this surprised a lot of people because, in fact, between 1990 and 2000, the economic characteristics of the community improved quite dramatically. In the east-side Detroit area we've studied, mean family income went up more than $7,000, and poverty rates went down about a quarter, but still death rates went up. That's something to ponder that I'll come back to.

But I took seriously that this session is meant to be about moving from theory to population study. And so, let me talk about the theory. About 20 years ago, I proposed an analytic framework which I called "weathering" where, in effect, I suggested that U.S. blacks and whites of the same chronological age might have different biological ages -- that 30 would not equal 30, and 50 would be even less likely to equal 50.

I say "in effect" because I originally came to this idea not just as a social scientist, but really more of a political theorist concerned with the wear and tear of



For details of such calculations, see Geronimus et al. *Journal of Epidemiology and Community Health* 1999; 53(6):325-334.

living life at society's margins, and not through considering the biology of aging at all. My qualitative work to that point with black residents of high-poverty areas, as well as large literatures that many of you are, I'm sure, familiar with in sociology and economics and anthropology, had documented that U.S. blacks were

more likely than their white counterparts to experience stressful situations, such as interpersonal discrimination, structural discrimination in housing and employment, material hardship, and multiple unpaid caregiving roles. I conceptualized weathering as a process reflecting the cumulative biological impact, both of having such experiences and of responding to or coping with, them. That is, I hypothesized that black Americans experienced early health deterioration as a consequence of their repeated experiences with social or economic adversity that called for sustained and high-effort coping and sustained cognitive and emotional engagement. And we all know now that such sustained coping can induce stress-related disease. At the time, people weren't talking about that, at least not where I was.

So if stress does fuel the progression of chronic disease, early health decline could be a physical price paid by blacks who work actively to mitigate, resist, or undo the ideological, economic, and social barriers to their achievement and wellbeing that are pervasive in our race-conscious society. And I want to highlight that in that sense, weathering is unlike more common theories to explain racial disparities in health that emphasize youthful indiscretion, individual health behavior, or material deprivation alone, because it allows for the possibility that premature health deterioration might not be alleviated, and, in fact, might be exacerbated by working hard, and fulfilling responsibilities. This is one of the more perverse consequences of our particular version of being a racially-stratified society.

It may also have implications for how we think about what's a mediator of these relationships. Because what I'm arguing, and others such as Sherman James have also argued and found empirical support for, is that in some structural locations, having, for instance, what you might call a strong internal locus of control may be hazardous to your health, rather than protective of your health. You can just think about it. If what your strong locus of control does is make you bat your head against the wall or glass ceiling and nothing more, that's not good for you. You have to not only have that strong locus of control, but also the resources that enable your efforts to have a positive impact.

So I'll leave that for the moment, and just summarize that the weathering hypothesis suggests the health of African-Americans begins to decline in early adulthood, deteriorates at an accelerated rate, and that it does so as a physical manifestation of the cumulative impact of repeated experience with material hardship, psychosocial challenges, and social exclusion. And the result could be, according to the hypothesis, that the stress inherent in living in a race-conscious society that stigmatizes and disadvantages blacks, may cause blacks to experience the morbidity and mortality of whites that are significantly older.

Mortality of Black Residents of Central City Detroit, Ages 15-64, 1990 and 2000

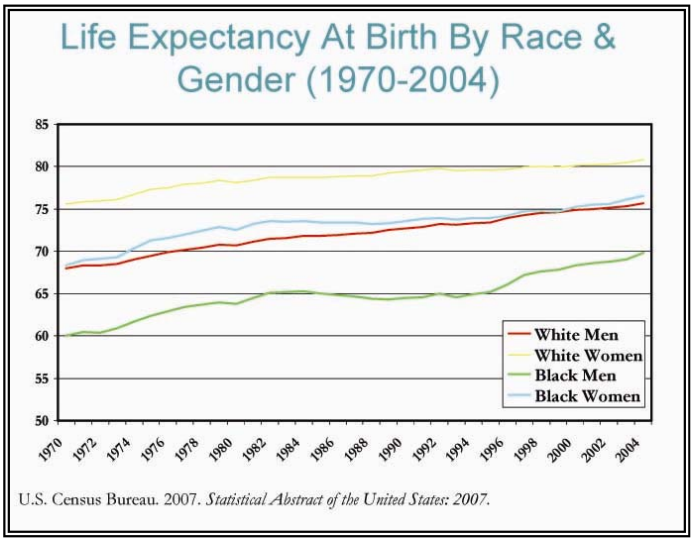
	Death Rate*		Excess Death Rate**	
	1990	2000	1990	2000
Men	**1163**	**1047**	**746**	**654**
Women	**580**	**641**	**355**	**407**
US Black Men	791	684	374	291
US Black Women	429	439	195	214

*Age-Standardized Death Rate
**Excess Death Rate relative to same sex/age whites. Rates are per 100,000 population.

The more ethnographic and qualitative work has provided some evidence of what might be everyday mechanisms for weathering. This is not an exhaustive list, and it's not meant to suggest a weighting of these factors, but just to give you some sense. It includes, of course, material hardship; cumulative exposure to environmental hazards and ambient or social stressors in residential and work environments; persistent psychosocial stress and high-effort coping, increasing in young to middle adulthood as family leadership roles are assumed and obligations expand and compete; increased pressure to adopt unhealthy behaviors; the early development of chronic conditions, themselves, can be stressful in a variety of ways; the increasingly deleterious impact of medical under-service; and the internalized effects of stigma or frustration, anger or rage at racial injustice.

And I mention these to show that weathering does encompass the usual suspects for explaining health disparities, namely behavior and income and medical service, but it does so in a particular way, and adds to them some of these more clearly psychosocial factors. In addition, different populations could be exposed to some of these factors more than others, and so you'll see -- and we have seen some evidence of --

weathering across, for instance, the socioeconomic spectrum, but in different degrees, and also across types of residential areas. So that, for example, the figures I showed you relating to the high-poverty



U.S. Census Bureau. 2007. *Statistical Abstract of the United States: 2007.*

urban areas are more dramatic than when we've looked in high-poverty rural areas, even when those rural populations are as, or more, poor than the urban areas.

I also want to highlight here that ethnographic studies and qualitative interviews suggest that young adult and middle-aged women, in particular, are central to extended-family economies and caretaking systems in poor black communities, and often must contend with competing obligations between work and family and dependent care, not just for children, but for ailing folks and elderly folks. And this, overall, I think is part of what Thom would call an "adaptive strategy," so I don't want to suggest it's a bad thing, but given that adaptive strategy, women of a certain age are particularly vulnerable to weathering.

"And what I'd like you to notice is that, for instance, in Chicago and Detroit, black men are as likely to die by age 45 as whites in the nation are to die by age 65."

For example, Linda Burton has found in her study of primary caregivers in three high-poverty urban areas, that these women talked about the ways in which, as one participant put it, they could "never get a break," which can be seen as an indication of chronic stress. And she also found that although she had a very young sample -- 83% of her primary caregivers were younger than 39 -- that 60% of them already suffered multiple morbidities.

To make this a little more vivid, I'll give you the example of Beverly, who was interviewed in by Helen Epstein in Yonkers, New York, in a high-poverty area. Epstein noted that "Beverly has asthma, diabetes, high blood pressure, rheumatoid arthritis, gout, and an enlarged heart. She is 48. And she had her first heart attack in her late 20s. One of her brothers died of heart failure at 50, and another died of kidney failure at 45, as did a sister who was 35. A young cousin recently died of cancer. In the past three years, at least eleven young people she knows have died, most of them not from gunshot wounds or drug overdoses, but from disease."

"About 20 years ago, I proposed an analytic framework which I called "weathering" where, in effect, I suggested that U.S. blacks and whites of the same chronological age might have different biological ages -- that 30 would not equal 30, and 50 would be even less likely to equal 50."

This may sound very extreme, but it actually is not so extreme, and in the interview studies I've done, I've found similar circumstances reported with some frequency. It does suggest that whatever quantitative

Probability of Survival to Ages 45, 65, & 85 Conditional on Survival to Age 16 , Men in Selected Populations, 2000*

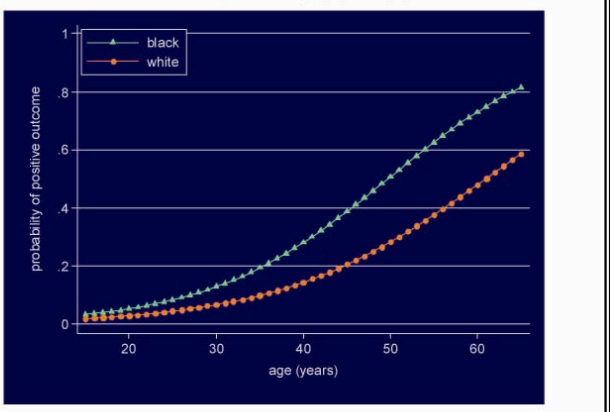
Survival to Age:	U.S. Whites	U.S. Blacks	Southside Chicago‡	Eastside Detroit ‡	Watts LA ‡
45	.95	.92	.82	.83	.90
65	.80	.68	.49	.54	.62
85	.27	.16	.09	.13	.12

*Mortality calculations based on data from the 2000 Census (adjusted for coverage error) and from death certificates for 1999-2001. ‡ entire nonHispanic Black population(~ 100,000) residing in high poverty census tracts
Source: Geronimus, Bound and Colen 2007.

study we do of a single health indicator the results are likely to be tip-of-the-iceberg, or pale in comparison to the enormity of the issues.

I'm not going to talk about this here, but for those who are interested, there are obviously clinical and also policy implications of weathering.

Let me turn to a little evidence related to biomeasures. I'm going to talk briefly about hypertension prevalence, allostatic load, and telomere length, the latter being very preliminary work that I hope to be able to expand. In terms of age prevalence of hypertension, everybody here is certainly aware that U.S. blacks have higher hypertension rates than U.S. whites. But I want to highlight an age dimension to those disparities. The weathering hypothesis would predict that the age trajectories of hypertension prevalence through middle age would be steeper for black than white Americans, with a wider racial gap in hypertension prevalence in middle age than at younger ages.



Geronimus et al. *Ethnicity and Disease* 2007, 17 (1): 40-48.

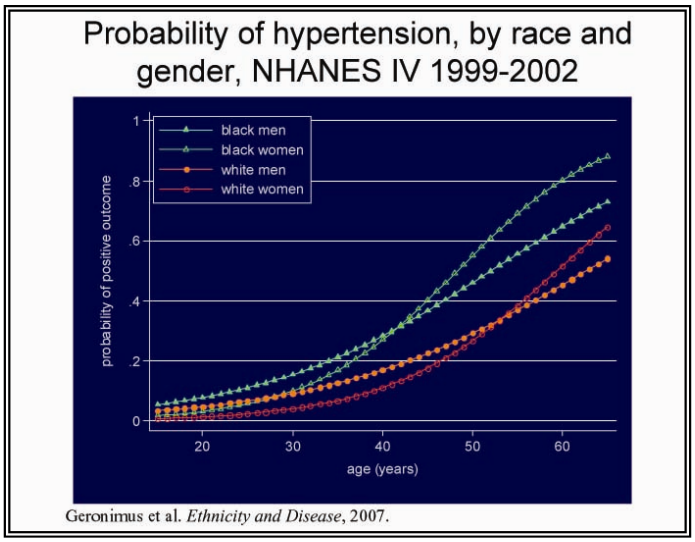
And I won't go through all the details of the study methods. We used NHANES-4. I can give anyone the paper who wants it. We used logistic regression. We adjusted for poverty and for BMI. We accounted for the complex sample design and did a range of robustness tests.

But the basic findings are these: You might call this a weathering age pattern, where the black-white differences start out fairly small, but then enlarge with age. And I want to point out that if we break it down further by gender as well as race, that the disparity among women is bigger than the disparity among men, and that, in fact, there are crossovers whereby black women's hypertension prevalence is higher than white men's. It crosses over somewhere in the 20s. Then it crosses over black men's around 40, to become the highest prevalence of hypertension of these groups. And white women's hypertension prevalence also crosses over white men's, but not until around the menopausal transition. But it does, then, begin to approach that of black men.

We looked, comparing NHANES-3 to NHANES-4 data, which are about ten years apart, the first being in the late '80s, early '90s; the second in the late '90s, early 2000s. And you don't understand, but this is actually two curves, but for white men they stayed identical over this 10-year period, so it looks like one curve.

For black men, the differences were also small. But for women, white and black, there were statistically significant differences by survey wave, for black women beginning around age 35, and for white women age 45, whereby their age trajectory of hypertension got steeper over that 10-year period. Here's the white women -- comparing the two survey waves -- and here's the black.

I wanted to point that out because...to say, first of all, these problems seem to be getting worse. They are getting worse just as we've all been told to be working to make them better, so we're doing something wrong, clearly -- or certainly not whatever would be right. It also, for those of you who are concerned about issues of cohort effects, suggests that if there are cohort effects, they suggest that the age trajectories using the NHANES-4 data understate the degree to which the age gradient in black women's hypertension prevalence has increased, or increases with age.

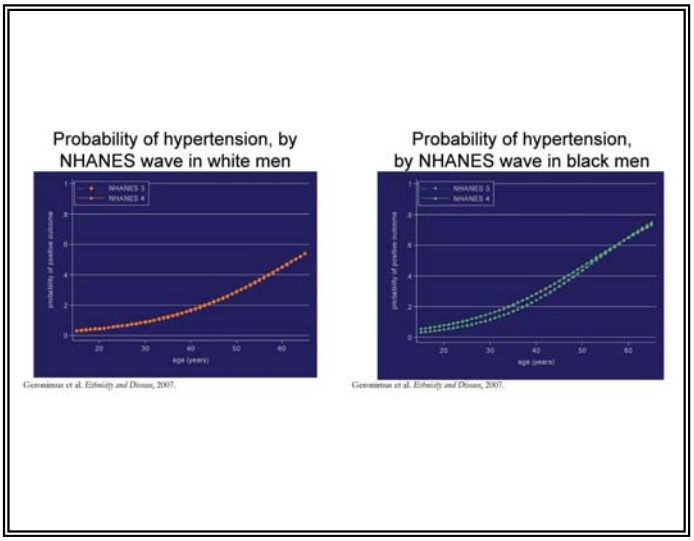


Geronimus et al. *Ethnicity and Disease*, 2007.

So then allostatic load, I don't need to tell you about how it is measured, thanks to Noreen.

The weathering hypothesis predicts that blacks have

higher allostatic load than whites, and the black-white disparity will increase in young adulthood through middle age. Again, we use NHANES-4. I think there are all the problems with the allostatic load measure



that Noreen mentioned, and to add further, in this case, with NHANES-4, we have none of the primary mediators.

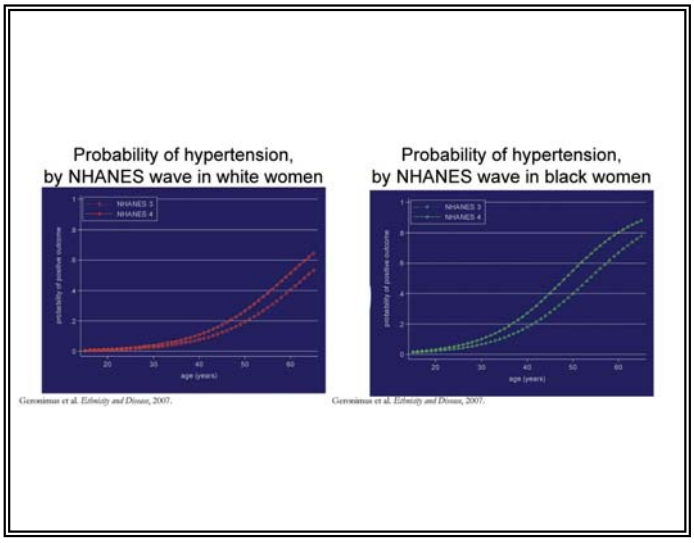
I really want to highlight that -- that this is all based on the so-called secondary mediators. But if you want to look at racial disparities and starting with people of younger ages, at this point, I think this what you have to look at. Hopefully, that will change. We did it very similarly to the hypertension analyses in terms of methods, and got similar findings, as you can see.

I'll go through these quickly. Black women have higher allostatic load score, or are more likely to have a high allostatic load score at every age. There's not a crossover. They just always have the higher scores in this case. And, in fact, here white women and white men are virtually identical -- statistically, they're identical throughout the ages. So the hypertension and allostatic load results are consistent, broadly speaking, with the weathering hypothesis. The black women appear to have the steepest age gradient increase, and their age gradient increase seems to have gotten steeper in the mid- to late-1990s.

So then, turning to telomeres, I'll just say that the cellular mechanisms for weathering are not, obviously, determined. But it does seem like telomere length might be a good way to consider this, and although she hasn't spoken yet, I'm going to leave it to Elissa Epel to tell you much more about that.

But what I do want to point out from Elissa Epel's previous studies, and part of what caught our eye to consider telomere length as a possible summary measure of weathering, was that in some of her work, she had found evidence that stress is related to accelerated cellular aging in young and middle-aged women. So it was, sort of, the right -- well not totally the "right" demographic, because it was a totally white demographic -- but in terms of age, it was about the right ages.

And also, Epel found that these markers were associated with increased perceived stress in their sample, their entire sample, but also with length of caregiving in women who were caring for a chronically ill child. That their study was conducted in young to middle-aged women, and pertained to some of the stressors that are routinely faced by black women in the U.S., suggested to us that maybe telomere length might provide insight into the cellular process of weathering. If telomere length *is* responsive to biological stress activation, and if weathering is produced by persistent high-effort coping with stressors, then blacks would be expected to have shorter mean telomere length through middle adulthood than same-age whites; or I might call that being biologically older than whites of the same chronological age.



Now, a decent study of this waits to be done, and we've certainly tried to develop a proposal to do one in Detroit. But as an initial step, and with some funding from NIA through the Michigan Center for the Demography of Aging, we've done a little bit of pilot work, analyzing a sample of respondents from the Study of Women's Health Across the Nation, or

SWAN for those of you who are familiar with it, which is a multi-site longitudinal study designed to examine the health of women in their middle years.

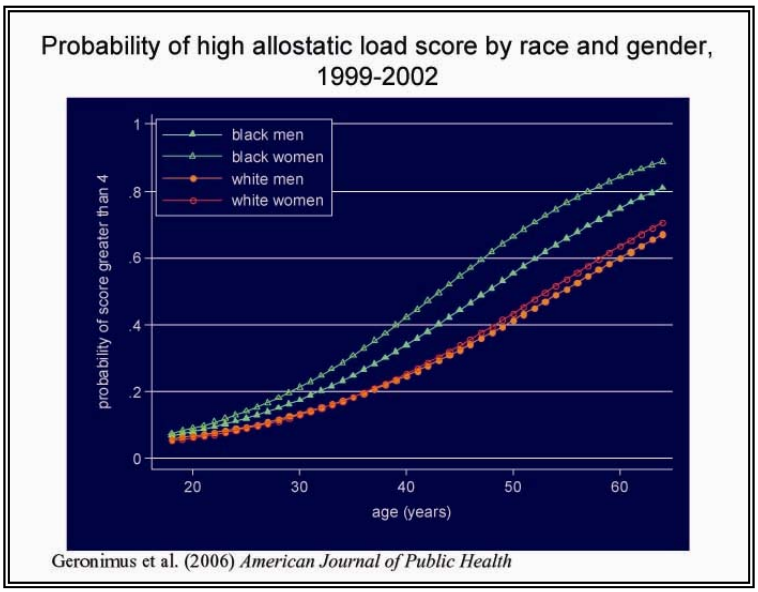
And I want to highlight just how problematic a choice that is for this particular purpose, but it was what we were able to do with the resources we had. The really key limitation is that the respondents may be too old for looking at what we want to see, because they are 49 to 55 years old, and menopause, itself, can trigger telomere shortening. If there were differences at younger ages between black and white women in mean telomere length, those may have been diluted by the time we get to women in their 50s. So it would have been nicer to have, say, women in their late-30s and, maybe earlier 40s, to study this, and we hope to be able to do that in our future work.

The sample is small from the selected sites, but most importantly, there's kind of an extra noise added, because the DNA -- the cells that we used -- were immortalized by Epstein-Barr virus. Those of you who know about that, know that that EBV immortalization affects telomere length, itself. Everything I've learned about how this immortalization approach affects telomere length is that it introduces random variation. There's no reason to believe it's systematic with relation with race, so it's not a completely lethal threat to the study, but it sure made it harder.

I want to highlight that all of these limitations would increase the chance of Type-2 error, and you'll see that we have no results that are statistically significant at conventional levels, although the p-values in some cases are p<.15. But just to give you an idea of our suggestive results, we estimated, when we controlled just for FSH to deal with the impact of estrogen on telomere length, that blacks, were 8.4 years biologically older than the same chronologically-aged whites; and that when we controlled for education as well, our estimates suggest blacks to be about six years biologically older than same-age whites.

Then the results get even wilder, so I'm not quite sure whether I should even be presenting these, but just take them with a pillar of salt -- that blacks are estimated to be almost thirteen years biologically older than same-age whites when we control for income-to-needs averaged over seven years, and for marital status. When we add in waist-hip ratio, menopause, smoking, vitamin usage, that the black women were estimated to be twenty years older than the same-age whites.

And again, please don't take any of these literally, but I think what *is* suggestive about them is they all go in the hypothesized direction, and the first couple are not *that* far off from conventional levels of statistical significance. We hope this at least help makes the case for expanding our work.



We're hoping to do a primary data collection effort in Detroit, where we'll actually have information on a very broad range of stressors -- environmental, built environment stressors, pollution, noise; as well as perceived stress -- perceived stress due to discrimination, perceived stress due to other things, life event stress. Also food intake food environment, access to healthy foods, physical activity, etc. Just a very broad range where we'd be closer to testing different steps of weathering directly.

For the moment, I would just say that the hypertension and allostatic load findings to date do provide some evidence that the impact of chronic stress on health has important implications not only for individuals, but also at the population level, and that they suggest ways that dynamic social relations between racial and ethnic groups may shape health in a race-conscious society. I think they suggest that progress in understanding and eliminating racial health inequality requires attention to how the interaction of American race and gender heuristics exact a physical price across a range of biological systems, for black Americans who engage and cope with stressful life conditions.

I also want to thank some of my collaborators on a variety of this work, including John Bound and Maggie Hicken, who's here, and Cindy Colen and many others, as well as for funding from NIH, including NIA through the Michigan Center on the Demography of Aging, from CDC, and from the Robert Wood Johnson Foundation, among others. And thank you.

[*Applause.*]

Q & A with Dr. Geronimus

Q: Black people, you said, on average, very roughly speaking, are twelve to twenty years older biologically.

Geronimus: Women. This was just the women in midlife. And the findings are just suggestive at this point.

Q: Did they have higher prevalence of any diseases that are associated with late age?

Geronimus: In the SWAN data, per se,... I'm not sure... I don't think we looked specifically at diseases in those data. It's just very well documented, and documented by us through our work with NHANES, as well as the work of others, that Black women have higher rates of chronic disease at earlier ages than whites. There's just consistent evidence that there are higher rates of hypertension, diabetes, a range of other diseases.

Q: So, one of my questions about the SWAN cohort, in particular, is one of the things we know, looking at black-white differences... And I'm assuming this was just Michigan site-specific, right? All four sites?

Geronimus: If we had any hope of having any power, we had to include women from multiple SWAN sites.

Q: So we know that they're heavier, so they have higher BMIs, higher rates of hypertension, more cardiovascular risk factors across the board. And on all of the stress measures in SWAN -- the CESD, the Perceived Stress Scale, the Everyday Discrimination Scale, negative life events -- the African-American women are consistently higher than the Caucasian women in SWAN. Even if you look within educational levels -- because SWAN has a fairly high percentage of college-educated women, both African-American and Caucasian -- and I'm wondering if you've had an opportunity to look at the telomere length after adjusting for some of those factors, and if you're still seeing...

Geronimus: Well, we adjusted for socioeconomic characteristics and for some of the anthropometric factors …and we intended to also adjust for some of those stress variables. But we had already...just putting in some of the few things we had in, already, exceeded our power to study this. So it just seemed that putting in the stressor variables, too, was just going to make things too thin.

We've also worried about the immortalization of the cells. As I said, as best I understand, that introduces random variation, but there have certainly been people who have reacted by saying, "You just can't use those cells at all." I think they're incorrect – as best I've followed this -- and this is where I think social science is sometimes helpful, to be able to understand the differences between random variation and systematic variation. But at the same time, they're right in that I think we need very large samples to overcome that random variation problem, so that we could look at these other things. I think there's only so far you can go with SWAN in terms of studying telomere length, which is unfortunate, because it has so many strengths.

Q: Actually, I had one other question. I'm sorry. My research looks at psychosocial factors and health outcomes in African-American women. And I'm wondering if some of the data that you presented showing that these effects were particularly pronounced…the disparities were particularly pronounced for the black women. Do you think that the traditional stress measures that we use -- perceived stress, life events -- are really capturing the life experience of.

Geronimus: Absolutely not.

Q: Okay. So that would be my thought, as well.

Geronimus: Part of what we're hoping to do in our Detroit study is have a broader set of measured stressors that are more specific to the population. Some of the stress measures there were developed through focus groups and other ways with members of the actual population. And then, as I said, we also have a lot of these environmental stress measures that people would not perceive, necessarily, or even be

aware of, but might still be affecting them physiologically.

Q: Thank you.

Q: Thank you. That is very interesting and, I think, important analysis. You show a lot of compelling black-white differences. But we know there's a lot of heterogeneity in African American communities in the United States, as well Caucasian communities. And complex interactions among race, ethnicity, class, and health. So have you looked at that? Tried to look at these... Are there different patterns of weathering by class within race?

Geronimus: Yes. We have looked. Well, first of all, the original slides on residential areas, as I said, were part of bigger studies with many more residential areas, that we selected to vary in terms of poverty rates, rural-urban, a range of things, and we found different degrees of weathering. It's very hard, and NHANES. We love NHANES because it not only has these clinical measures, and some socio-demographic ones, but because it also includes the young adult through middle ages, which I think is where the action happens here. Many available data sets have only old folks or only kids. But at the same time, there aren't enough blacks in NHANES to, then, break it down by some of these factors, so that's a limitation there. So the slides I showed you, we weren't able to look at all the sociodemographic factors we would have liked to.

But what we have found is that any way we've sliced it, there's evidence of weathering among African-Americans, at least for women. And it's just the degree. And when you think about it -- and I think this is one of the problems with thinking more about socio-economic status than about structural location or race/ethnicity -- is that you can think about economically better-off African-Americans, and women will report token stress, and role overload, and even greater rage with racism and interpersonal discrimination than African Americans who are economically worse off.

And so when I listed those so-called everyday mechanisms for weathering, you can kind of think about where you are -- whether it's in terms of class, or geography, or gender -- which of those might apply, and then how many, and with what intensity or duration.

Q: That was very interesting. My question is along similar lines. Your argument about weathering is that it's a form of persistent high-effort coping. Most of what you've been describing has to do with exposure rather than with coping. And I'm just wondering the extent to which the variations in that type of coping happen actually to relate to the biomarkers and the endpoints you've been studying.

Geronimus: Okay. Well, I think that the choice of allostatic load, and if telomere length reflects biological stress activation, that choice also reflects an interest in high-effort coping. I know I went pretty fast, but we had controlled for poverty in these. We had actually looked at some obesity measures. And those factors don't make weathering go away. So, it's also in some ways, theoretically, a sense of what could be going on if it's not just material inputs, or it's not the usual suspects related to behavior. So the biomeasures were chosen theoretically as ones that are thought to be stress mediated.
Hypertension is a stress-related disease, allostatic load is construed as stress-mediated wear and tear, and telomere shortening appears to be accelerated by stress. Using these theoretically chosen measures, we were inferring the linkage to stress as a possible explanation for the age gradient dimesnsion of racial disparities. It's sort of a strategy that you may or may not like, of, well, this is how I'm theorizing the world works, and if it works like this, then you should expect to find this-with-this-outcome, or that-with-that-outcome. But we don't have data that really let us do all those steps to really, really see whether that's the chain.

Q: So just to clarify, you're inferring high-effort coping from the pattern of the dependent variable, rather than as a predictor.

Geronimus: No, I'm inferring high-effort coping more from other kinds of research that vividly and consistently describes the high-effort coping that African-Americans persistently engage in. Then, thinking about *if* that would impact their health, this is how I might think it would. But certainly what we had hoped to do with SWAN , and what we still hope to do with the Detroit primary data collection study is actually have all of those measures, so that we can look and see whether the world really seems to be working that way, step by step.

Q: I just have a quick comment, which is that the

method for assessing telomere length may also be controversial -- that the PCR method has been questioned, and there are some other methods which are much more labor-intensive and time-consuming, but seem to have a better specificity, and a better validity and reproducibility. So I think that that may be another thing that has to be considered. And I just wanted to bring that up in this forum since people may be interested in measuring telomere length. I also think that immortalized cells is not the place to measure telomere length.

Geronimus: You're in very good company thinking that. As I said, I think that view is questionable, but it's certainly not what we would choose to do. And it probably shouldn't be done if for no other reason than enough people *think* it shouldn't be done, that whether they're completely right or not, you're going to have a harder time , probably, publishing or securing funding doing it because it is subobtimal. But I would love to hear anything from anyone related to telomere length measurements.

Q: So I wanted to make a suggestion that would improve your p- values. You showed us results for level effects -- black-white differences -- but as you pointed out earlier in the talk, the hypothesis is that the differences should get bigger with time. So you should be estimating a model that has a level, race, interaction with age in the telomere analysis. And that would dramatically improve your p-values.

Geronimus: Yeah, the age is so compressed in the SWAN data, there's just very little to do... all of the respondents are within a six-year age range that crosses the menopausal transition And I wouldn't think at that six-year age interval, the differences would get bigger over time, because of menopause. But your general point is well taken, and we've actually, with some of the other work, done exactly that. Thank you.

Q: Arline, I think your work is very clear on middle age -- the influence, the departure, the disparity in middle age -- but I'm less clear about what's happening in later life. So if I look at your telomere -- your preliminary results, it would almost imply a shrinking of the gap, or perhaps a stability of the racial differences. So, I hear you loud and clear -- weathering occurs in middle age. Very clear. What's happening in later life?

Geronimus: A variety of things. First of all, if you think about the first slides on mortality, a lot of blacks are dead by old age. So you just have a selection going on there. Also, people whether black or white start developing some of these diseases just *as* they age, so the white folks as a group are getting worse off. So I think it's just very hard to think about those ages. The numbers get also thin. We've tried to look at them, and the numbers start getting really tiny. Just think about how if 50% to two-thirds of high-poverty black men are dead before they turn 65, it's kind of hard to look at them at older ages, and to think about who is it you still have there, as well as the issues of numbers. So, probably given the selection issues, among older folks, the disparities are not as large due to survivor bias. If you could do weathering through the whole age span, it would probably get bigger in middle age, and then appear to start converging at older ages for reasons of selection and the prevalence of chronic disease at the older ages.

Short Presentations: ***Biomarkers of Stress and Aging Cortisol and Beyond***

Emma Adam, PhD (Northwestern)

Today I'm going to talk about a paper that I coauthored with John Cacioppo, Louise Hawkley and Brigitte Kudielka. This paper reflects my general interest in studying how everyday experiences in naturalistic settings get under the skin to influence HPA-axis activity, as indicated by cortisol levels; and also looking at the implications of changes in cortisol for everyday functioning, as well as emotional and physical health.

I introduced this quote two years ago when I spoke, and this was a quote from a reviewer on a career award: "Cortisol will break you heart." And I appreciated the sagely advice of this reviewer. The reviewer did go on to award me the grant, but with this warning regarding the dire consequences of focusing on cortisol in my career.

"Dysregulation, I argue, is not whatever cortisol pattern you find in your less healthy group... I would argue that dysregulation is a failure of cortisol levels to appropriately calibrate themselves to the changing demands of both momentary experience, as well as the changing demands of daily experience."

When I came to this conference two years ago, the initial title given to my talk was: "Biomarkers Gone Bad: The Case of Cortisol." I think I was supposed to fly in as the cortisol defender to assure everybody that it's not that cortisol doesn't make sense; it's just that we haven't yet made good sense *of* it. I argued that the reason that we haven't made good sense of it yet is that we don't typically measure it very well.

So the basic premise for today -- I'm not going to repeat the messages that I gave last time, but rather extend my points regarding cortisol measurement, and really focus on the idea that a cortisol level, one cortisol level, has little meaning because cortisol is part of a stress-responsive system. The fact that cortisol levels are designed to increase in response to stress implies that it's *intended* to be dynamic. And so if we're going to understand the role of cortisol in health and disease processes, we need to move beyond cross-sectional analyses to start to look at cortisol dynamics; we need to look at change over time in cortisol levels.

We do know that cortisol changes on the order of moments -- we call this 'cortisol reactivity'. It also changes within the day. That's the diurnal cortisol rhythm. We also suspect, although very few people actually study this in a longitudinal manner in humans, that cortisol levels and diurnal rhythms change over the order of weeks, months and years in adaptation to chronic stress.

We haven't done much longitudinal analysis with humans because we don't like to wait weeks, months and years to study change over time. For many health outcomes, you do need longitudinal data over the course of years, but I'm going to argue you can also get important clues from short-term change data. While you're impatiently waiting for change-across-years data, you can look at day-to-day changes in cortisol and experience to get some important insights into the functioning of this system.

Typically, when researchers gather multiple days of cortisol data, they average across days to extract the meaningful trait variation and get rid of that annoying state in error variation. But what if that state variation is actually important -- that day-to-day state variation? What if day-to-day calibration of cortisol levels in response to -- or perhaps in anticipation of -- changing experience is an important element of HPA axis function?

My hypothesis driving this study and some future work I'll be doing is that some of the day-to-day variability that we see in cortisol often considered either error or uninteresting state variation is actually systematic and functional variation in response to changing experience. And I would argue that day-to-day calibration of cortisol levels in response to changing experience is actually an important indicator of healthy HPA axis functioning. So if you're not changing levels from day to day in response to the demands of the day, that is perhaps a sign of HPA-axis dysregulation.

The sample I used for this study was the Chicago Health Aging and Social Relations Study. Most of my research is actually on adolescent populations, but I had the pleasure of having a chance to collaborate with John Cacioppo and his colleagues on this data set looking at cortisol data in older adults. The sample for my particular analyses was 156 older-aged adults aged

50 to 67. The sample as a whole was selected to be representative of individuals in that age range in Cook County, Illinois. There was also some oversampling for blacks and Hispanics and an attempt to keep an equal gender distribution in the sample.

The measures that I'll use are three days of data on daily experiences and cortisol. The daily experiences were gathered using an evening diary method, so people were reflecting on their experiences during that specific day each evening for three days. They also gathered salivary cortisol levels at wake-up, thirty minutes post to wakening, and at bedtime on each of the same three days. From these three cortisol measures, we got the cortisol parameters, or aspects of cortisol functioning of interest were the wake-up levels, the size of the cortisol awakening response, and the slope of the diurnal cortisol rhythm. I'll describe those in a little more detail in a minute.

We used MEMS Track Caps ® to electronically monitor compliance of the timing of the cortisol samples so we had some way of calibrating whether or not they took the samples when they said they did, based on use of this electronic device that has a timestamp associated with it. We also had questionnaire measures of health problems, health behaviors, medication use, stress, social support, and psychological well-being.

The daily experience data I'll focus on today involves the mood state data with a few other things that aren't quite mood thrown in here. We factor analyzed the list of adjectives on which participants rated their daily experiences and six dimensions of daily experience emerged: 1) sadness/loneliness, including things like lonely, sad, isolated, discouraged, helpless, intimidated, threatened; 2) active/effective, which involves feeling lively, energetic and good about oneself; 3) tense/ angry, including tense, angry, uneasy, annoyed; 4) symptomatic/fatigued, which involves feeling physical symptoms and fatigue; 5) confident, which involved feeling confident, like things are going their way, and 6) confused/forgetful.

Interestingly, it was the negative aspects of experience (sadness/loneliness, tense/angry, and symptomatic/ fatigued) that were most strongly associated with cortisol in this study, so I'll be focusing on those. There were no significant results for the other dimensions.

The parameters of cortisol functioning I focus on are the wake-up value, the slope of the diurnal curve from wakeup to bedtime, and the size of the cortisol awakening response. The latter, the size of the awakening response, you could calculate from a difference from the wakeup value to the wakeup plus 30 minute value, but I actually calculate it as a deviation from the curve that defines the slope. So I look at the slope from wake-up to bedtime, the morning value, and the size of the awakening response, calculated as a deviation of the 30 minutes post-awakening value from the curve at that time of day.

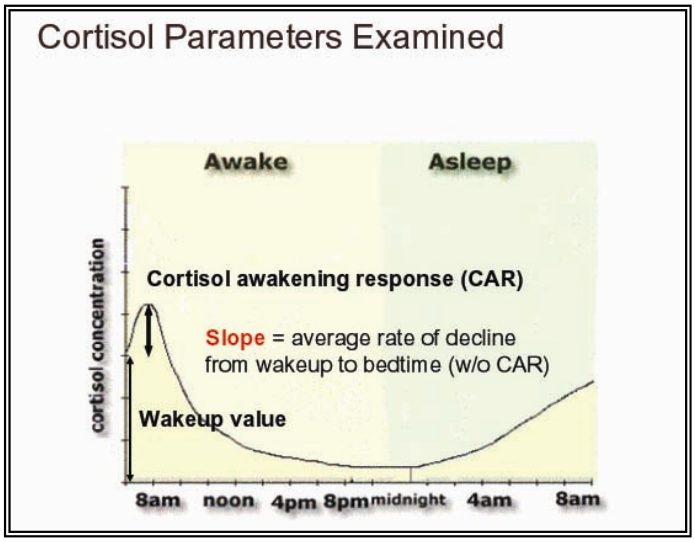
We can talk about the interpretation of these three different parameters, perhaps, in the discussion. The interpretations are subject to debate, but what's important and what I noted the last time I spoke -- that these different elements or parameters of cortisol functioning appear to be both predicted by different experiences and predictive *of* different health outcomes.

"The fact that cortisol levels are designed to increase in response to stress implies that it's intended to be dynamic. And so if we're going to understand the role of cortisol in health and disease processes, we need to move beyond cross-sectional analyses to start to look at cortisol dynamics; we need to look at change over time in cortisol levels."

These are just examples individuals' diurnal rhythms -- this is actually not from the study, so there are more data points here, but these graphs illustrate exactly how different diurnal cortisol rhythms can be for different individuals. This person here has a relatively low morning cortisol level, and a whoppingly huge cortisol awakening response. The slope, if it were defined from here to here, is relatively flat. The next person has no cortisol awakening response but a relatively flat curve, also. So there's large differences in these parameters between individuals, and what I'll talk about today is that there's also changes in these rhythms within individuals from day to day, and you can predict the changes in these parameters from day to day systematically from changes in experience.

For my analysis, I used a three-level, multilevel growth model. I'm not going to go into huge detail, but I'll give a few details for those of you who are really geeky

about these kinds of things. Level 1 is where I modeled the diurnal rhythm, including the wake-up level, the changing cortisol across the day, which is the diurnal slope, and the size of the awakening response.



At Level 2, I represented daily experience -- changes in experience from day to day. And Level 3 is where I modeled person-level characteristics -- individual differences that are stable -- as well as daily experiences that are averaged or combined across the three days of testing.

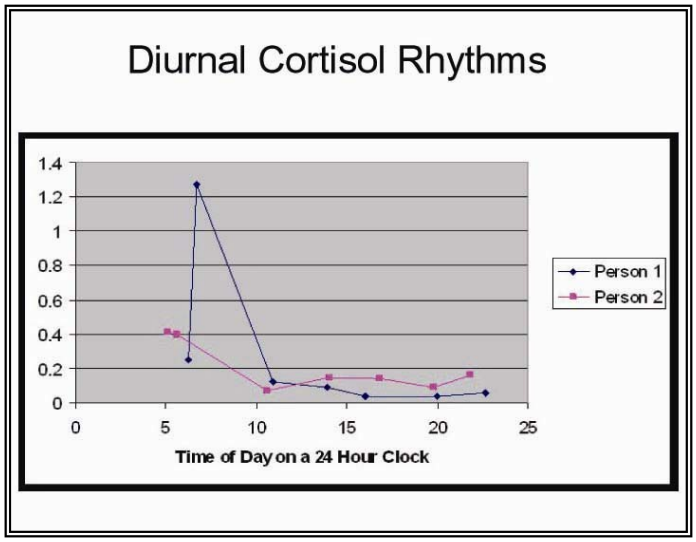
We don't need to go into the details of this modeling, but if anybody's interested, we can talk about those, except to note that there were multiple control variables in the analysis. At the person level we controlled for demographic variables, mental and physical health and medication use, at the day level we controlled for time of waking and hours of sleep the night before. We also looked at sleep quality the night before. Finally, also at the day level, we controlled for whether or not they took their samples on time each day, with a day-level dummy variable indicating whether they were compliant with the timing of sampling that day according to the electronic monitoring data.

Five minutes! Whoa. I'll get quickly to the results. We first looked at a cross-sectional analysis, which is what most people would do with these data, which is just to associate average experience on the three days of testing with average cortisol parameters, and found that one standard deviation higher sadness or loneliness was associated with, on average, a thirteen percent higher cortisol awakening response. Anger/frustration was associated with a one percent flatter diurnal cortisol slope.

Finally, and perhaps most interestingly for this older adult sample, one standard deviation higher average fatigue was associated with a lower average wake-up cortisol level. There's actually an association with a lower cortisol awakening response, too, in the cross-sectional analyses. The general pattern is higher fatigue, lower morning cortisol.

The question is -- we've got these cross-sectional associations between cortisol and experience -- does the experience predict the cortisol, or does the cortisol predict the experience? What's the chicken and egg here? We can't answer that definitively without experimental data, but can start to try to get at with some longitudinal analyses, and change across days *is* longitudinal.

So the next step was to predict cortisol parameters on Day 2 and 3 simultaneously from both day-before and same-day experience, and basically see which one wins out as the better predictor of cortisol on a particular day. Does day-before experience predict cortisol better than cortisol that day predicts experience later that day? The answer is that it's a little bit of both. In terms of the association between loneliness and the cortisol awakening response, day-before loneliness predicts a higher cortisol awakening response the next morning with no association between cortisol



awakening responses that morning and sadness and loneliness that same day.

For anger/frustration, a flatter diurnal day-before anger/frustration didn't predict next-day diurnal cortisol slopes, but the diurnal cortisol slope that day is

associated with same-day anger/frustration. As it turns out, the direction actually is from anger/ frustration to the cortisol slope because it's the evening level that's defining this flatter slope. So basically, anger during the day predicts a higher evening level later that day, which accounts for a flatter slope that particular day.

"...some future work I'll be doing is that some of the day-to-day variability that we see in cortisol often considered either error or uninteresting state variation is actually systematic and functional variation in response to changing experience."

With respect to the fatigue, however, day-before fatigue has no relation to morning cortisol the next morning. But if you have high morning cortisol on a particular morning, you have lower fatigue for the rest of the day. Or put another way, if you have low morning cortisol on a particular morning of testing, you have higher fatigue for the rest of that same day. So there seems to be some evidence that both cortisol parameters on any particular day are affected by prior mood state, either the day before or earlier that day, but wake-up cortisol levels on a particular morning appear to affect fatigue levels for the rest of the day.

So conclusions are that looking at interrelated changes in experience and cortisol over time, even over short periods of time such as day to day, shed better light on cortisol's role in human health and functioning than looking at cross-sectional associations. And the results really illustrate the dual role of cortisol as both responsive to emotional experience, as we often look at it, but also as we know it to be -- an important regulatory hormone, being a contributor to the energetic state of the individual. So it's really a bi-directional influence with certain types of emotional experience activating cortisol, and then cortisol having effects on the individual's functioning throughout the day.

Just one more note regarding dysregulation, because it's a pet peeve of mine. Dysregulation, I argue, is not whatever cortisol pattern you find in your less healthy group. Often, researchers will have a clinical group and a normal group, and whatever cortisol pattern they find in their clinical group they'll call dysregulated, which is a little circular. I would argue that dysregulation is a failure of cortisol levels to appropriately calibrate themselves to the changing demands of both momentary experience, as well as the changing demands of daily experience. And this can be, as Elissa Epel pointed out, either by under- or overshooting responses to momentary and daily challenge.

What's really novel about this particular paper is that I'm introducing the notion that there's a day-to-day calibration as well as a moment-to-moment calibration of cortisol levels to experience that's important for cortisol's role in human functioning. The reference to this particular paper is available here, if you're interested in reading all the details. I want to thank my co-authors, the participants, the National Institute of Aging, and our other funders for their important roles in this particular project. Thank you.

[*Applause.*]

Q & A with Dr. Adam

Q: Thank you. That was a very nice integration over a period of time. I wonder if you thought about the fact of chronic stress over a longer period of time, causing, really, adrenal suppression or what we call adrenal deficiency syndrome. Dr. Epel was talking about the DEXA scan that patients, rather than hypertrophying their muscles from too much exercise, the adrenal tires out and is unable, then, to respond in stressful situations. So there's a relative adrenal insufficiency, and that could be reflected kind of on a larger scale than the day-to-day feelings.

Adam: Absolutely. And so in the individual difference level, you would expect people that long term stress is reflected in changes in typical or average levels of cortisol. But I actually would expect it to be reflected in day-to-day variation, as well. So those who have this sort of adrenal exhaustion, some form of burnout, would actually no longer have this effective day-to-day calibration of the morning levels and the cortisol awakening response. My argument and interpretation of the awakening response is that it's actually the...it's a response to the anticipated demands of that particular day. It's kind of like your extra cup of coffee in the morning when you know you've got a challenging day ahead. So someone who no longer has that day-to-day calibration of their morning cortisol, getting an extra

boost when needed, is not going to be functioning as effectively. And that's something that may have developed over time through chronic stress.

Q: Right. And that's something we see clinically with patients with, say, chronic fatigue syndrome or other kinds of chronic -- you know, I just can't get going -- that may be really what one of the energetic factors is that you've identified. Thanks for your presentation. Very good.

Adam: Absolutely, yeah.

Q: Quick follow-up on that. Do you actually have in the data as to whether these were anticipated demands or whether they were surprising stressors that came along?

Adam: We have the type of stressors, but we don't have, the question did you anticipate this or not. We could code the qualitative data, but I'm not sure there's going to be enough there that would give us an indication of whether it's anticipated. I've done some very preliminary analyses with other data where I did ask about anticipated demands, and that was a strong predictor, but that was a very small data set. I think it's an important variable to add to studies that are looking at the cortisol awakening response.

Steptoe: Yeah. I just wanted to make a comment on that. The most obvious example of that are the studies comparing weekend and weekday.

Adam: Absolutely, yes. And that's...

Steptoe: Because it has a consistently lower cortisol awakening response on weekend days, which is associated with lower demands, as people measure the mood of the day.

Adam: Right. And we get that weekend effect in these data, as well. Yeah.

Q: One thing that I'm wondering about here... I mean, it makes sense to me that higher cortisol in the morning, lower fatigue that day, because this is mobilizing resources and so on, like you're saying.

Adam: Right.

Q: And I think that aspect of it is really interesting, because it shows that more is actually better in this case. It's not just more is bad.

Adam: Right.

Q: But how do you interpret the functional effect of the higher overnight levels, which is driving some of this flatter slope that you're talking about? Have you put any thought into that? What is the body trying to do by keeping its cortisol levels kind of elevated and not dropping down as much overnight?

Adam: Well, I don't have data on what's happening overnight. If you're talking about the evening levels before bedtime, that means that the person is still coping with challenges throughout the day. So I don't think that's anticipated. I think the evening levels are responsive to the daily experiences.

And then the question is, how effectively do those levels clear and shut down overnight and affect values the next morning? I have done a day to day analysis of whether it's as simple as having high evening levels the night before predicts lower wake-up values the next morning. It doesn't seem to be that straightforward – levels the night before do not seem to predict either higher or lower levels the next morning. It's not just a simple carryover or negative feedback thing.

Q: How did you control for medication use? Was it a dichotomous -- any medication use or not, or specific medications?

Adam: We looked at specific medications to the extent that we had them. And we looked at specific medical conditions, as well. And there were some significant associations that I'm not reporting here. Alcohol use was associated with a much higher awakening response. Having a psychiatric disorder was associated with a higher awakening response. So there were, in addition, things that I didn't present here about various physical health and medication effects.

In this older adult population, it's really hard to throw people out of the sample as so many people have medical conditions of one form or another or are on medications. Certainly, if someone had an endocrine disorder or were taking steroid based medications, they would be eliminated from the sample. But for the most part, to the extent possible, we statistically controlled. We also statistically controlled for noncompliance, which helped the estimation quite a bit. People who took their samples late had a lower

cortisol awakening response, so we had noncompliance as a dummy in the model as well.

[*Applause*.]

Reference

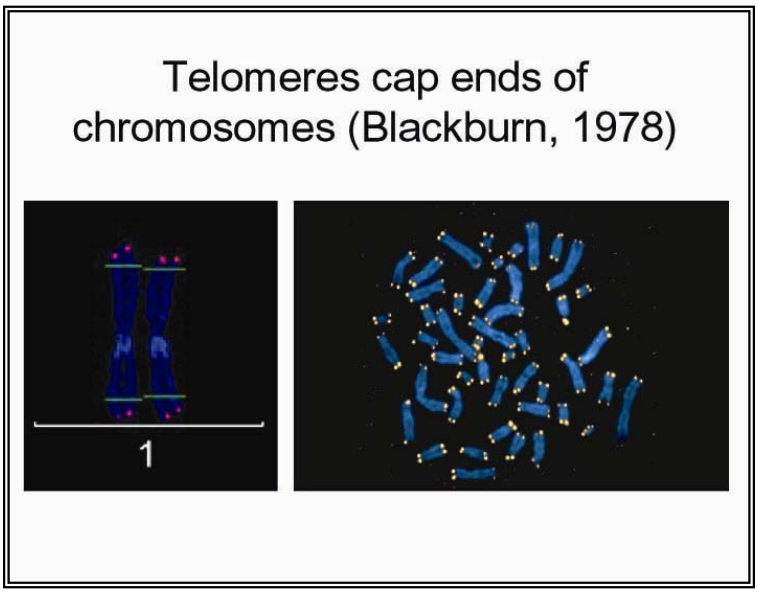
Adam, E. K., Hawkley, L. C., Kudielka, B. M. & Cacioppo, J. T. (2006). Day-to-day dynamics of experience-cortisol associations in a population-based sample of older adults. *Proceedings of the National Academy of Sciences, 103*, 17058-17063.

Elissa Epel, PhD (UCSF)

I wanted to start by thanking you inviting me -- Stacy and Lisbeth. This is a really great nexus of researchers in these different areas with a common goal and interest.

"A biomarker is going to be related to age; and it is going to change with age."

Before I start, I want to acknowledge my cell biologist collaborators. Elizabeth Blackburn, who discovered telomeres and telomerase over twenty years ago, and Dr. Jue Lin, a post-doc we're both supervising, who has focused on understanding stress pathways to cell aging. So she's doing a lot of the in-vitro basic biology work alongside Liz and I, who are trying to understand what's happening in people.



I share a common driving question about health disparities, fundamental causes, and the social environment and stress pathways, like many of you here. So I'm going to start with this question: "How does social class get under the skin, and is it through chronic stress?"

So this question started from these classic early stress researchers, Canon, Selye, and Geronimus' weathering hypothesis and a recent conceptualization of allostatic load. The latter two have more overlap than divergence -- this idea that chronic wear and tear on physiology is going to lead to premature aging. That's where I really want to focus: aging, not disease.

So what is happening early on in their twenties? What is starting to break down? What is biological aging in the absence of disease? What makes skin wrinkle and hair turn grey? Are there common processes of disease progression? So if we're really interested in a biomarker, a preclinical disease marker, we don't want to operationalize aging as early-disease onset. We really want to know what's happening to healthy people in mid-life or earlier.

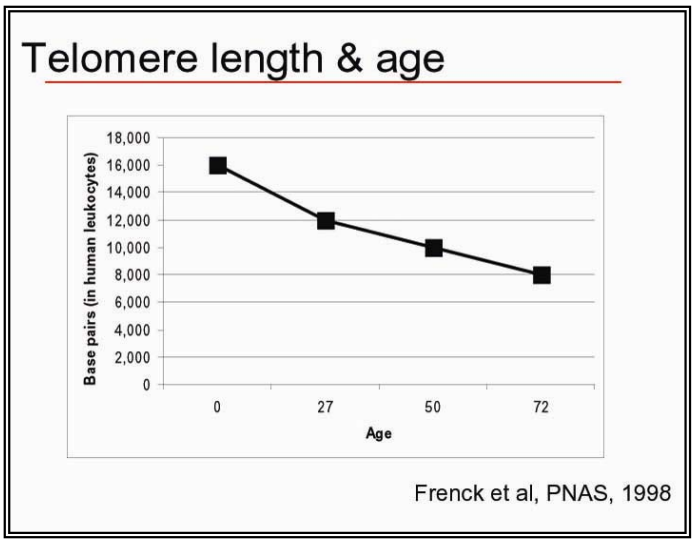
Again, traditional disease risk factors are probably a little bit too late to study, because there's already ongoing disease processes. But is there a measure of biological age that might change before disease and serve as the cellular basis of development of disease. So biomarkers certainly can bring us closer to these. But we have to remember that biomarkers are also a reflection of pathology, so in older people, they might be reflecting ongoing disease rather than biological age.

A biomarker is going to be related to age; and it is going to change with age. The other definition of a biomarker is that it must predict longevity. And ideally, the biomarker is not just a marker, an epi-phenomena or marker, but actually involved in the mechanism of early aging, in the pathophysiology of aging -- or I won't pathologize it--in the physiology of aging. So a biostress marker is a marker that is malleable by environmental exposures and stress, and I think that is a common interest to this group. What are biostress markers that are measurable?

And, of course, as you know from my title, I'm going to be focusing on telomere length. I've been interested in a marker that we can examine during a person's lifespan, developmentally, that measures aging. I'm going to talk about cell aging with telomeres. I'm focusing on telomeres, how they related to health and stress, and then will talk about some methodological considerations in measuring them…All in ten minutes or so.

So these little lit-up chunks of DNA are the telomeres. These are chromosomes. And chromosomes are capped by these DNA chunks that protect the chromosomes from fusing and breaking. So they're not the genetic material. They play an important role in health. Every time these mitotic cells divide, the telomeres shorten. And with progressive cell division,

and increased shortening, they get to a terminal length and they don't shorten anymore. When they reach a terminal length, the cell tends to go into senescence. It doesn't necessarily die or go into apoptosis, rather, it is no longer able to function and perform its duties.



So, for example, if we're talking about an immune cell, CD8 cell, a senescence CD8 cell has lost its surface markers, it can't recognize antigens, and it starts changing some of its functions. It starts secreting more proinflammatory cytokines, for example. So senescent cells don't do us very much good. In fact, they can do harm.

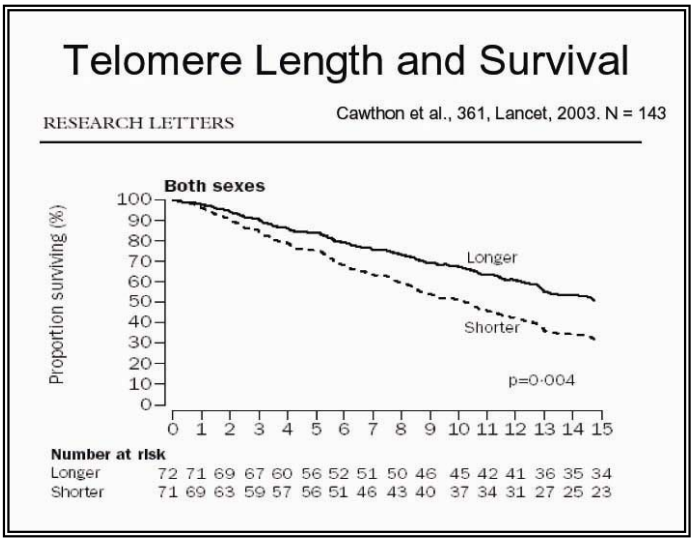
I'm going to be focusing on telomere length in white blood cells today. But I want to point out that most of the replenishing tissues in our bodies, the tissues that we need to divide and multiply to keep us healthy through the decades of life, are dependent upon these telomeres. This telomere/telomerase system is necessary to keep these cells dividing over a long lifetime. So this has implications, for example, for the cardiovascular lining. These cells are dependent on telomeres and telomerase. And in the immune system, if cells have short telomeres, our immune system can go into what has been called immunosenescence.

Telomeres do change with age. There's a lot of scatter in this relationship. The correlations may be .40 or so across studies. So you can see that as we age, we do have more mitotic division or clonal expansion of our immune system, and that leads to shorter telomeres. Cell division leads to shorter telomeres, but there are other factors, as well, that lead to shorter telomeres.

So for example, the biochemical environment can shorten telomeres. We don't know that much about what in the biochemical environment has these effects. We do know that oxidative stress is one of the factors that can damage and shorten telomeres. Cytokines can also stimulate cells to turn over and thus, telomeres will shorten. And so telomere length is partly a reflection of many different things. It is not a pure marker of stress. I'll be talking about relationships with stress, but clearly, it's a marker that's multifactorially determined; it's partly genetically determined, it's determined by lifetime exposure to antigens, as well as the biochemical environment.

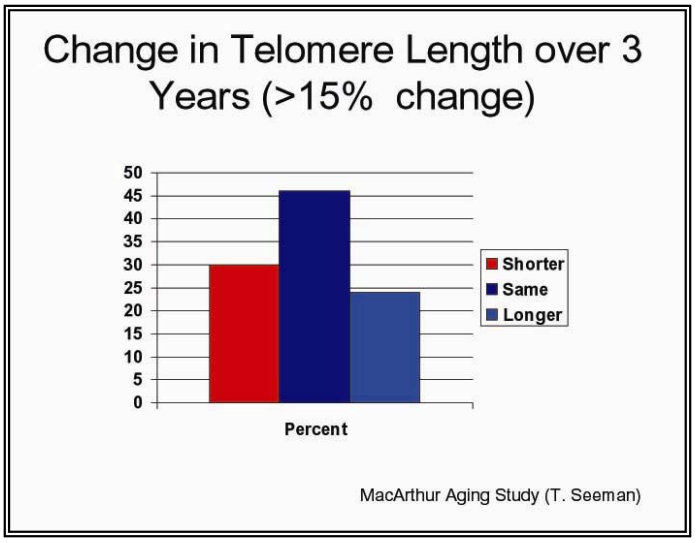
So this has been studied for the last twenty years, in in-vitro systems, mainly, such as yeast. And in the last five or so years, there has been an explosion of research in humans because it's become easy to measure in humans. Because it's a predictor of longevity in the cell, it's been assumed to be a predictor of longevity in humans. We now know that across many studies cross-sectionally, telomere length is associated with a wide variety of diseases. So shorter telomere length in leukocytes is associated with, for example, cardiovascular disease, heart attacks, diabetes, vascular dementia, Alzheimer's disease, as well as rheumatoid arthritis and lower bone density and osteoporosis.

Most of these studies are cross-sectional, so it does beg the question whether telomere length is really just a



marker of disease status, a predictor, or actively involved in the mechanism of these degenerative diseases, these chronic diseases. The landmark study by Richard Cawthon in 2003 found what many had predicted, which is that telomere length in mid-life predicted mortality. The people with shorter telomere

lengths tended to die four years earlier for men. The difference is almost five years earlier for women. So telomere length in mid-life predicted earlier mortality.



We have recently replicated these results in Teresa Seeman's MacArthur Healthy Aging cohort. And we found a similar effect size for women. Just as a note, you can see that both of our samples were under 200.

So one question is, if short telomeres can predict mortality, is this a malleable factor? Can telomere length be modulated? In the MacArthur Aging Study, we looked at telomere change over a three-year period. This hasn't been looked at in this short of a period. There's only been one study that looked at telomere length change over time. It was around a ten-year period. So if telomere length can get longer, it can either be through cells, the immune system having a shift in the cell populations that are there or making new cells, or there could be a per-cell increase in telomere length.

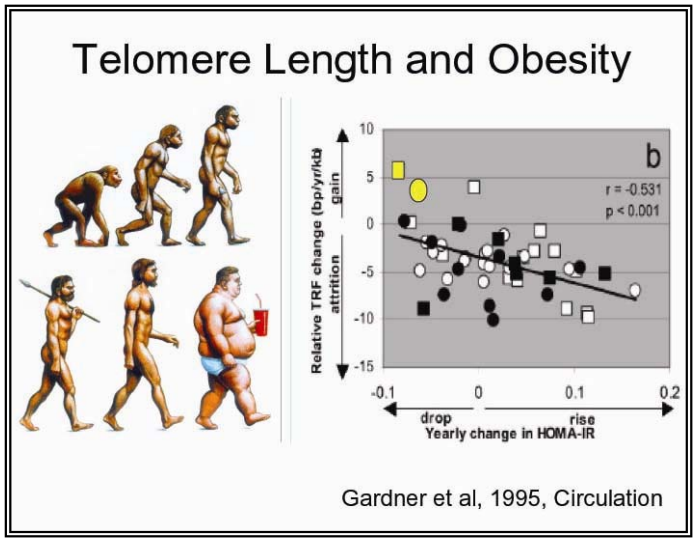
And how would that happen? There is an enzyme, telomerase, that actually increases telomere length in a preexisting cell, so it actually adds back base pairs. I won't be talking about telomerase much today because it's more difficult to measure, but it is a driver of telomere length. So we're extremely interested in telomerase.

You can see from the MacArthur Aging Study that people's telomere length decreased as predicted -- about thirty percent -- but most maintained telomere length, and some increased. So we're very interested in finding out what can cause increases in telomere length. In this longitudinal study that I mentioned, they found that obesity was causing a decrease in telomere length over time; we're really in trouble because we have an obese society -- mostly an overweight society, but largely obese, as well.

You can see that the two people here who show the biggest increases in telomere length over time were two who lost dramatic amounts of weight over this ten-year period. So clearly, weight and insulin resistance are co-varying with changes in telomere length.

Is cell aging related to chronic stress? That's been the question I started out with and have been most interested in. So we looked at two samples of healthy women with ranging stress levels. This has already been published, so I'm going to go through this quickly. Healthy young women showed a correlation with telomere length. We're now looking at healthy older women, post-menopausal women. Some dementia caregivers, some healthy controls appear in blue. They have the same relationship -- a linear gradient with perceived stress and telomere length. Here you can see that exposure to years of caregiving is associated with shorter telomere length.

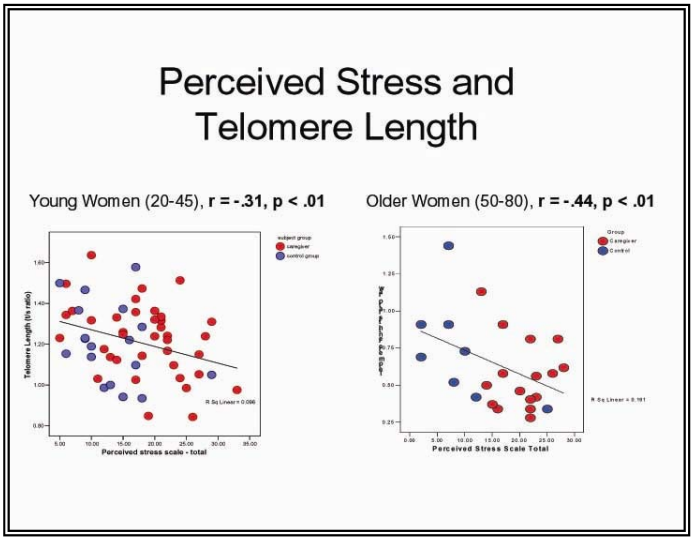
So in addition to our finding and replication with perceived stress, others have now found this extension to related disorders like major depression. People with major depression have shorter telomere length than



controls. An experimental study which stresses out mice has found that the stressed mice developed shorter telomeres. We've also found correlations with nocturnal stress hormones, cortisol and catecholamines, but that by no means indicates that it is a causal agent. That was just a correlation...although

we do think that stress arousal might be involved in the pathway.

So if, at an individual level, chronic stress and stress perception are driving shortened telomeres, does the



forest shape the trees? Does stress exposure by neighborhood, by culture, or by environment shorten telomeres? One study looked at social status, and found that blue-collar had shorter telomeres than white-collar female workers. No one has looked at neighborhood, race, ethnicity, or culture much at all. There's a lot in the works with some of these large studies. I added culture because I learned yesterday we could operationalize culture in sophisticated ways. Thank you, Thom (McDade).

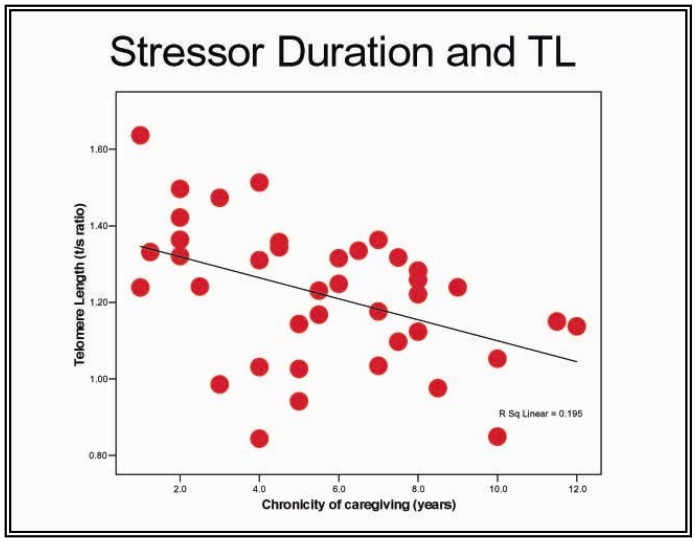
So some quick words about measurement of telomeres. Southern blots are the most expensive methods. They are several hundred dollars. They require a lot of DNA. They are extremely reliable. This is the gold standard. It's a direct measure using a gel. PCR is a newer method developed by Richard Cawthon, and the benefits are, it's closer to fifteen dollars a measure. It's very quick. You can do high throughput, and it doesn't measure subtelomeric repeats, like the Southern Blot does.

And I'll just talk about some caveats. You really have to know what you're doing there. It's not an assay anyone can pick up based on a recipe and do well. You have to control the co-variation between assays. And then another method -- the fish flow method -- requires whole cells, so it's limited in that way for use in epidemiological studies.

In terms of sample size, I'm going to refer you to a paper that nicely lays out estimated sample sizes needed to look at telomere length change. It was written by Abraham Aviv. The estimates are purely based on a twin study in the UK. I, as a caveat, tell you not to be discouraged by these estimates. They say you need several thousand people to look at any factor, but do remember that the studies I showed you on mortality used several hundred people, and they did find effects. The UK study had a large age range, which adds variance.

If you're looking at small samples, then you can reduce your variance by controlling big, confounding factors. For example, don't include smokers. Smoking has a large effect on telomere length. It's probably going to outweigh any psychological effect or other smaller effects. So in controlled studies, we select eligible people out of hundreds of people, matching on BMI, age, measuring health behaviors, eliminating people with certain medications and health conditions, because these can all influence telomere length. So in that way, if you have a highly controlled study, you can reveal larger effects in telomere length.

Assay variability is really the biggest problem here. We're trying to measure differences in base pairs, and so you really need to have the highest quality control possible... I have several colleagues who have started up the telomere length PCR by someone in their medical school who did not have experience with this



assay. Their results were null, and we don't know if they had true null results or real null results. So you really want, to compare results, compare with samples in a gold standard lab, and test inter/intra assay variability throughout batches, and have the appropriate controls throughout batches.

This is basic to most lab people, but these are really things to take seriously. Test samples in triplicate. Test samples as much as you can using the same lots of the reagents. So there really are a lot of considerations to make sure that, for example, a PCR assay is going to be valid across the lifetime of your study.

"And so telomere length is partly a reflection of many different things. It is not a pure marker of stress. I'll be talking about relationships with stress, but clearly, it's a marker that's multifactorially determined; it's partly genetically determined, it's determined by lifetime exposure to antigens, as well as the biochemical environment."

I'm just going to end with some of the most interesting questions, to me, in the field right now. So: unanswered questions. There are a lot. I was going to start off by encouraging people to add this to their study, but I see that I don't need to. It's already being added to several studies. So genetic, epigenetic, and prenatal factors. We know there's a large genetic transmission. Is there also telomere shortening effects from maternal depression, maternal nutrition and maternal stress, factors that might also influence the size of telomeres in infancy?

Early childhood is a time of dramatic telomere shortening. The biggest difference in telomere life throughout the life span may have happened in early childhood. And so what we end up with in older life is partly determined by the first five years in life. We do lose base pairs every year, and that may be very important to health, but it's also important to look early on and see what's happening with transmission of telomere length and shortening in childhood. That has not been looked at very much.

How early can weathering of telomere length be observed? Can the shortening be reversed? We have several clinical trials using, for example, antidepressants and stress reduction to see if we can increase telomere length. Is telomere length a marker or a mechanism of premature aging? We simply don't know, and population studies won't be able to answer that. But in the meantime, they will be able to answer some of the important, big phenotypic questions—such as, "Does telomere length differ by race and ethnicity?"

[*Applause.*]

Q & A with Dr. Epel

Q: Correct me if I'm wrong -- that all these studies you've done were done in women? Are there different processes in women versus men, or is that just the samples you happened to have?

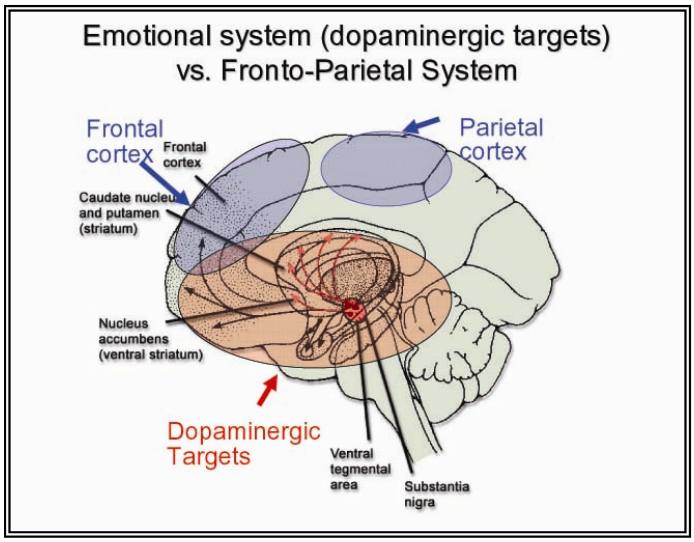
Epel: That's a good question. The two stress studies I showed you were in women. I do small controlled clinical studies. All of the health studies, the probably twenty studies linking telomere to health, were done mostly in men, some in women, and the mortality study was done in men and women. We don't know if there are gender differences. Women tend to have longer telomeres. Telomeres are regulated in part by estrogen. There's probably a big shift -- a drop in telomere length during the **perimenopausal** and menopausal period. No one has looked at that period specifically.

Thank you.

[*Applause.*]

David Laibson, PhD (Harvard)

First, thanks very much for inviting us to join you today. I'm thrilled to be here and I've learned a great deal. I'm an economist, so I come to this with a very different perspective. I'm actually two steps behind David Weir in learning all these new things. I want to introduce my collaborators. Daniel Benjamin, who is here today, is also an economist, though Dan really straddles economics and psychology. Chris Chabris is a psychologist, and Edward Glaeser is an economist. And I'll identify some more collaborators as we go through some of the research.



I want to first introduce a little bit of intellectual history. Economics is in the midst of a potential revolution. Economists historically believed that everyone optimized and everyone basically was rational. In the last twenty years, there's been a sea change. The new view, in essence, is that psychology and biology and neuroscience all matter for economic analysis, and there's a very active debate within economics about the potential role of these new disciplines.

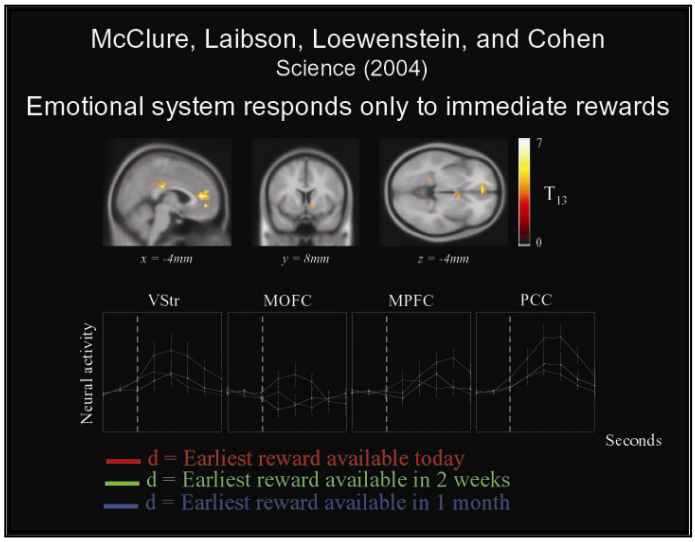
It's ironic that our conversation today is happening here at the University of Chicago, because if you had to point to an institution that was most aligned with the traditional orthodox view of the rational actor model, it would be Chicago.

Economists are in the midst of a discussion that is changing the face of our field. The view of behavioral economists, or people who do psychological economics, is that we've got to take psychology and biology seriously, and if we do, economists will learn a lot more about human behavior. Among economists, there's a new interest in the neural foundations of decision-making and preferences, and I'll talk a bit about that.

Now I want to show you just an example of the kind of research that's taking place.

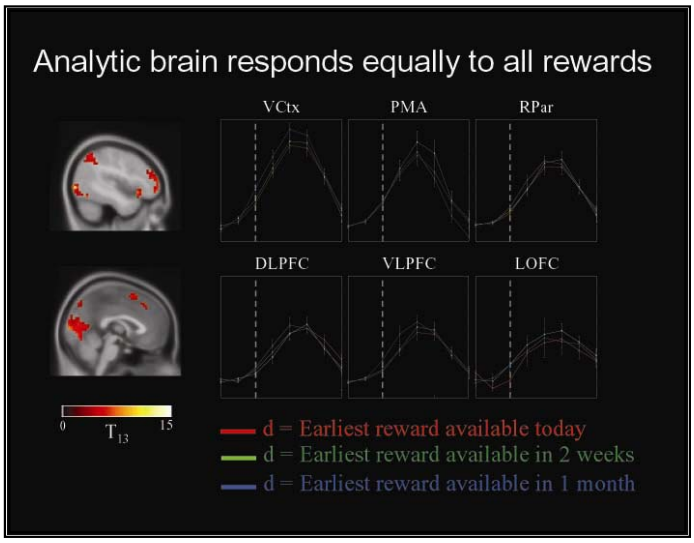
One hypothesis that my collaborators and I have been exploring is that there is an internal debate within the brain, between the impatient limbic system – more accurately the dopaminergic reward system -- and the patient, forwarding-looking fronto-parietal cortex. This internal conflict may produce a lot of the seemingly irrational or paradoxical behavior that flies in the face of the predictions of the classical economic model. This is joint work with Sam McClure, Keith Ericson, George Loewenstein and Jonathan Cohen.

Using neuroimaging methods we have measured the neural responses of these different brain regions. When we give our subjects a choice that involves an immediate option, the dopaminergic reward system becomes highly active. But when we give our subjects only delayed options to choose between, we see relatively little response in the dopaminergic reward system. We're measuring a "gift bias" in this emotional system.



By contrast, in the analytic systems -- in the parietal cortex and in the prefrontal cortex --we see almost no differentiation between immediate and delayed reward.

Research like this is beginning to get economists interested in unpacking the brain and considering the biological foundations of behavior. The next frontier in this work is to begin to unpack the genetic



influences on economic behavior. This is a research program that is about to begin -- we're not there yet.

As you know, when you compare MZ and DZ phenotypic correlations, there's a gap. The MZ correlations tend to be higher than the DZ correlations, and under certain very strong assumptions, one can infer heritability of a phenotype by doubling the MZ vs. DZ difference. Whether or not we believe all those identifying assumptions is a critical question which I'll leave for another day.

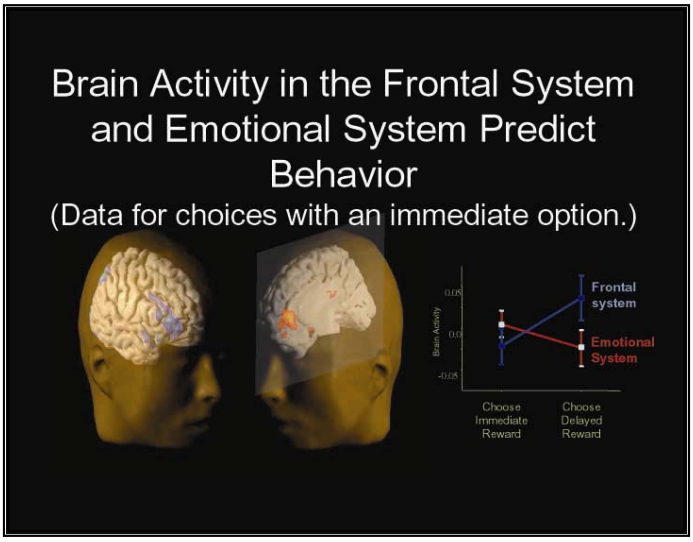
What's interesting for the purposes of today's talk is that when you look at economic phenotypes you see the same magnitude of gaps between MZ and DZ correlations. Whether we're looking at log wages, years of schooling, or log wealth -- work that's coming out of our collaboration with Kaare Christensen and the Longitudinal Study of Aging Danish Twins -- we see similar imputed levels of heritability. So it looks as though health behaviors are potentially as heritable as economic phenotypes, like wages, schooling, even wealth. It's particularly striking that wealth is generating a heritability level like that given that wealth is determined jointly by both the heads of the household.

The next step will take this kind of work into the realm of molecular genetics. At the time that the Human Genome Project was concluding, I think we were all very excited about what would happen next. Though it has happened less quickly than many of us anticipated, there are now many studies that deliver on that initial promise. On this slide I've given you a taste -- a small sampling -- of what's been happening.

There have been very prominent studies for coronary heart disease, diabetes, and seven common diseases that the Welcome Trust Case Control Consortium studied. These results have been reported in *Science* and *Nature*. There have also been replications. And there is the KIRBA allele analysis, which has also been replicated.

There's the older analysis by Caspi et al. -- gene environment interactions linking early life adverse experiences and the MAOA allele (the combination of the environmental risk factor and the allele risk factor produce super-additive vulnerabilities). More and more studies successfully link health outcomes and other psychological outcomes to genetic variation.

This slide summarizes another example of a very promising body of work. Here's a gene environment interaction that Caspi, et al. reported in 2003 linking a serotonin transporter gene to depression, interacting that effect with the number of adverse life events. We're seeing more and more studies like this; successfully linking outcomes to genetic variation at the molecular level.

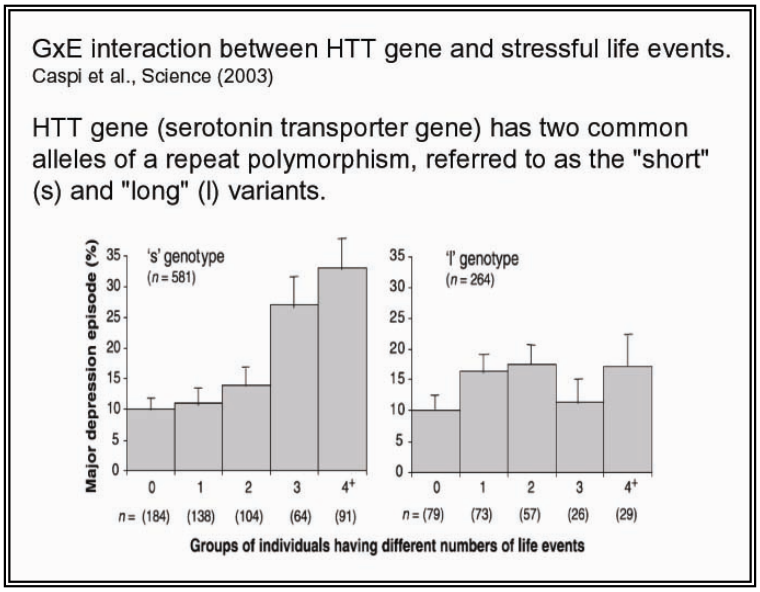


So why would economists be interested and excited by these possibilities? Economists do four things, and I want to argue that all the four things that we do are fruitfully advanced by these new genetic opportunities. I'll talk about explaining variation. I'll talk about measuring behavioral responses to incentives. I'll talk about helping to design public policy. I'll finally talk

about helping individuals make better choices. And I'll emphasize that some of these are still science fiction, but you can see where we're heading in the next 15-20 years.

First: explaining variation. First, there are obviously going to be some direct gene effects. In addition, we're likely to find gene-environment interactions. The complexity that comes from the gene-environment interaction is likely to be both a source of richness and a hurdle. We hope to identify new mechanisms. So, for example, when you do a whole genome analysis, you will often discover genetic variation that you hadn't hypothesized to be critical in explaining the particular source of variation. So if economists want to identify new pathways, whole genome analysis may actually be a very effective way of identifying pathways that we hadn't thought about.

Of course validating existing pathways with candidate gene studies is another natural way for us to proceed. We want to identify causal relationships. How do we do that? Well, we can think about genes as Mendelian randomization in certain cases, and we can study whether that particular pathway, using, say, family studies, actually generates the hypothesized effects, studying causality in cases like that.



And finally, even if we don't care about the genes, per se, as we accumulate a larger and larger body of knowledge about the linkages between genetic variation and behavioral variation, we can put those genes on the right hand side of our regressions and control for some of that noise, thereby enabling us to better infer relationships between the variables that we do care about on the right hand side.

Second: measuring behavioral responses to incentives. Ultimately what economists really get passionate about is understanding how variation in our environments, and the incentives that we face affects our behavior. The environment can be thought of as an incentive and we really want to understand how people respond to different environments, and how those responses may be mediated by their particular genetic starting point.

We're interested, fundamentally, in the gene-environment interactions. How do different people respond to differences in their environments? Are some people very receptive to the availability of alcohol, maybe so receptive that they drink in a self-defeating way? Are some people highly responsive to alcohol taxes, and hence easily controlled by changing the tax system? Are others perhaps unresponsive to alcohol taxes and as a consequence better helped by other policies like product bans or restrictions on access?

We want to understand all of these different pathways, and of course there are going to be individual differences in these environmental pathways, and studying those differences through the genetic lens, we think, can be very useful.

Third: we also help design policy, and that's been a big part of economics for over a century. Understanding biological mechanisms will be critical there as well. We want to target policies. We're already targeting policies based on individual groups that have the greatest vulnerabilities that perhaps may be best helped by that policy, or maybe are suffering the most in our social system. And to the extent that genetic vulnerabilities can help us better target those policies, we're obviously interested in using this filter.

Now that might sound kind of like a scary Brave New World, but of course we're already doing that. I mean, PKU is precisely a targeted intervention based on genetic vulnerability. So that could naturally be extended to a broader set of domains. It's already happening in the health area; it's going to probably happen in the economic area in the next few decades.

Fourth: we also want to help individuals make better decisions. Now at the moment -- and I think the

conversation yesterday about ethics was very telling -- we're not ready yet to tell people much about their genetic data, their personal genetic data, their personal genome, because that information is, as yet, not very predictive. But you can imagine a day, again, a decade or two from now, when knowing a lot about your personal genome enables you to make much better choices over your life course.

For example, macular degeneration is increasingly something that is predictable using genetic information. If I tell you, at age 25, there's a very good chance you're going to lose sight in mid-life, perhaps you will want to prepare for that. You may want to economically prepare for that. You may want to begin to accumulate resources so if you do need to stop work because of disability, you have the economic resources available to enable you to do that.

So we're interested in giving people potentially useful information that's going to help them make better economic choices. Now, obviously there's enormous ethical issues and great concern about giving people information they can use appropriately as opposed to information that's going to scare them, or information that basically has no predictive power. So these are all subtleties that we've got to worry about here.

Now, there are a lot of methodological hurdles. We don't yet know what the key phenotypes are that will be predicted by genetic variation. We don't know the level of aggregation. Should we be studying a broad phenomenon like wealth accumulation or a very, very precise phenomenon like discounting between rewards I get today and rewards I might get tomorrow?

What are the key environmental factors? What should we be measuring? Early life experiences? Are there other environmental factors that are critical sources of interaction? We don't know what to measure, and so we've got to keep an open mind.

Of course, there's all sorts of very substantial multiple testing problems. We all believe that gene environment interactions and gene-gene interactions are going to be important. Identifying these interactions will be a challenge. Given all of the necessary statistical analyses, we've got to worry about whether we are identifying real effects or false positives.

Now there have been, of course, many, many false positives in this literature in the last three or four years in the health domain. Obesity is one example. Herbert, et al. published a positive result in *Science* in 2006 and Luce, et al. published a rebuttal in 2006 in *Science* as well arguing that the original result did not replicate. And there have been hundreds of examples like that, prominent findings in molecular genetics that have not replicated. We're obviously highly sensitive to those concerns.

"Economics is in the midst of a potential revolution. Economists historically believed that everyone optimized and everyone basically was rational. In the last twenty years, there's been a sea change. The new view, in essence, is that psychology and biology and neuroscience all matter for economic analysis…"

How can we find polymorphisms with very small effects? We would guess that most phenotypes are driven by hundreds of genetic pathways and perhaps thousands of gene-environment interactions. Height, we now believe, is highly polygenic. There's a recent paper by Perola, et al. So identifying these very small effects is obviously a challenge, and we don't want to run after small effects if we can't actually measure them.

Is it possible that behavioral and social science phenotypes are more polygenic than health phenotypes? And if that's the case, perhaps we're wasting our time by trying to measure the social science phenotypes because they're just too small to pick up in the sample sizes that we have. Are we even looking in the right places? Perhaps we should be thinking about epigenetic mechanisms, or alternatively, copy number variants instead of standard snips. So there are many other alternatives that should perhaps be on our radar screen, and we've got to make some decisions early on about where we're going to look first. It's really an exciting opportunity, but one which poses many, many potential pitfalls.

There are obvious imperfect solutions to the problems that I've posed. None of these solutions really make us very comfortable, but they give us a bit of comfort. We want to now keep an open mind about the right phenotypes for social scientists to be measuring. We want to follow people through their entire life course. I was just talking with my colleagues who run the

Wisconsin Longitudinal Study, and excited to hear about all of the early life events that they're measuring.

We want to work with large samples, obviously. We want to be very focused on replication. We want to be very, very skeptical of our initial findings. We want to use priors. We want to test functional pathways. And perhaps we want to think about endophenotypes. I was enjoying Joe Lee's talk about the endophenotypes that he was using to validate some of his APOE analysis.

Now there are, of course, a slew of ethical issues, and I'm glad I don't have to actually repeat all of these challenging problems, because we had such a wonderful talk yesterday reviewing them. One thing that I might emphasize, though, that I think is quite encouraging is that, whereas in the health domain the ethical issues are, in many ways, overwhelming, in the economic domain there are a few outs (though the outs don't get you off the hook entirely). For instance, if you're simply observing that we now have found a polymorphism that influences, say, your willingness to hold growth stocks, you know what? You're not obligated to inform your subject that her genes make her slightly more likely to buy General Electric than General Motors.

Q: Aren't you obligated to tell their spouse?

[*Laughter.*]

Laibson: So in many ways the stakes are a lot lower in the social sciences because no one is going to die of being a little more prone to hold equities instead of bonds. So the ethical issues are very serious, but they may be a little less alarming in the domain of economic and social science behaviors.

Finally, there are cost benefit considerations, and I think here it's very, very exciting news. Basically, genotypic analysis is very expensive right now, but the costs are falling extraordinarily rapidly. They're not falling continuously, so you can go through three or four years at the same cost, and then all of a sudden a new technology comes out and the costs plummet by a factor of five. So what's happening over the long run is that costs are falling roughly by a factor of 100 every ten years, and I'm averaging, now, over the past decade.

And there are so many new companies that are basically in the market in this very competitive, exciting economic marketplace, offering better and better systems to provide genotypic analysis. So I think the cast of characters is familiar to many of you: Illumina, 454 Life Sciences, Affymetrix, Applied Biosystems, Helicos BioSciences and so on and so on. Every week, it seems to me, there's another announcement about someone else who's entering the fray. And of course there's George Church, who wants to do it all for free, basically, with his personal genome project. So there's a lot of opportunities here, particularly if you're willing to bank samples even for a few years.

I think the relevant comparison is to Moore's Law. As you all recall, this engineer forecast, in the '50s or '60s, I think it was the '50s, that transistors would advance such that you would be able to double the number of transistors on a chip every two years. Well, that would imply a factor, a kind of cost factor, of 32 every ten years.

Well, compare that to what we're experiencing in the chip domain, which is an improvement factor of 100 every ten years. We're proceeding much more quickly than we did in the extraordinary revolution of the transistor. The methodology is also changing, which is very exciting, so things like the Hap Map and whole genome scans and new inference techniques are also giving us a great deal of hope in terms of what is to come in genetic analysis in the next decade.

I think the sort-run strategy is to be skeptical, but also to recognize that there's enormous value in these analyses, probably, in the near future. I certainly would be encouraging surveys to collect and store DNA samples, even if they're not jumping in right now in the genotypic analysis. I think there's a lot of option value here, and we should be investing, making modest investments, seeing if there are any returns, and then following on with significant more investment if, in fact, we can demonstrate proof of concept studies.

So what are our ongoing studies? Even though we're skeptical, we are starting some exploratory studies. There are three studies that the group I told you about at the beginning is currently working on. We're collaborating with the Longitudinal Study of Aging Danish Twins, we're collaborating with AGES and the Reykjavik Study, and we're engaging in a new study which we're calling the Boston Study. Let me quickly

tell you about those three.

The first study is joint with Kaare Christensen and we're learning that a lot of economic phenotypes that are quite complex are actually heritable using standard behavioral genetic analysis – classical twin studies.

Our collaboration with AGES/Reykjavik Study involves Vilmundur Gudnason and Tammy Harris sitting here today. We're really excited about this. This is the first economic genotypic analysis using a panel of SNPs to study candidate genes. We are excited by the longitudinal structure of the AGES/Reykjavik data which goes back, now, 40 years. There's a lot of economic history in that panel, though it was originally designed for health purposes. We think this is a very promising place to begin. We're using a HapMap strategy plus a functional SNPs strategy to pick our SNPs.

And by the way, if you have any exciting SNPs that you think belong in this study, any particular functional or candidate SNPs that you think are interesting, this is the moment to tell us about them so we can include them in our panel. We're kind of solidifying our Illumina chip at the moment, so knowing, your suggestions in the next few weeks would be useful. In fact, in the next days would be really useful!

Female: [You need a] collaborative bent.

Laibson: Absolutely. [*Laughs.*] Fair enough, that's right. We economists just talk about everything. We're not worried; we don't have many secrets. But I take the point. Not everyone has our culture.

And finally, there's a new study that we're beginning to bring to the field, which we're calling the Boston Study. It is a whole genome analysis as opposed to the candidate gene analyses that I mentioned earlier. Here we're using the Affymetrix platform.

So we're very excited and we're looking for collaborators. We think that there is a lot to be done here as long as one proceeds carefully, skeptically, and with modest budgets. It may or may not work, but we think it's a risk worth taking. Thank you very much.

[*Applause.*]

Q & A with Dr. Laibson

Q: An interesting talk, but I must express a great deal of alarm about when economists get passionate about genetics, I certainly shudder.

[*Laughter.*]

Q: And the reason… I must give a self-disclosure now. I come from Red Clydeside and Red and Scotland bent, a Socialist background, and one of the conceptions we had when we were growing up is the big problem with advanced capitalist societies is they had to make some sort of an explanation why some people get a million times more reward than other people.

And this is happening even more and more in the United States, so the CEOs are getting now, what, is it 500 times the amount of money that the average worker makes, and it's getting more and more and more. So when you can't explain anything else any other way except this is a very, very, you know, a society with a lot of problems regarding equity, you say it's all genetics. And this has happened for 300, 400 years, ever since we thought about genetics.

And it reminds me of a little cartoon we used to use that was two little maggots sitting on the top of a cesspool, and one little maggot fell into the cesspool and the other little maggot didn't. And then about two months later, out comes the maggot again, who's seven times the size of the other maggot that didn't drop into the cesspool. And the maggot says, "How is it? What's your explanation?" He says, "It's all in the genes."

And it really bothers me that now you're taking economic outcomes using these genetic methods that… especially when you're looking at…even for diseases, Alzheimer's disease, that I know about, hypertension, there are multiple genetic models to understand risk.

You know, I understand projecting the relative weights of genetics versus environment, which they do in the Twin Study, but the problem with all that is you have to explain the context, because the relative weight of genetic versus environment is going to change dramatically, depending on the environment. You get to roses, exactly the same thing, exactly the same genetic material, give one water and one not, and one is going to die. The same genetics, but it's responsible to environment.

So I think you're entering into -- with enthusiasm, which is unfortunate, I hate to say it -- this minefield of an area with a considerable political background and a political history about this kind of approach to this political background, very naively. So I hate to tell you, and Tammy I'm surprised you let him, you know, you're so very… I would say, please, economists, stay out of genetics. [*Laughter*.]

Laibson: I don't disagree with any statement you made except the last one.

David Weir, Ph.D.

Thank you, Rena. Thank you, Stacy and Tom, for organizing yet another outstanding conference on this important topic. All right, I give Thom credit for way too much. And thanks to you, the hardy survivors, for sticking around for this last session.

I want to take just a moment to deconstruct the title that David and I were given. So first, "debate." I thought well, so who should we aspire to emulate in our debate? So Paul Ehrlich and Julian Simon was a sort of famous debate from my early schooling, and they, of course, wagered on a twenty-year projection of relative prices into the future, and I think Julian Simon actually won the bet. But we couldn't think of anything that we'd want to predict twenty years from now, so that didn't work. Something less profound than that, well, Cheney-Edwards, but I thought I would suffer in that analogy, so I passed on that one.

So the real problem is that David and I don't really disagree about very much, so debate's not really what we had in mind, so the analogy that came to me was the old Miller Lite Beer commercials in which Billy Martin or some other famous alcoholic sports figures would be brought together into a bar to debate the merits of Miller Lite. And one side would say, "Tastes great!" The other side would say, "Less filling!" And they would end up in a brawl, and finally in jail. And so this really seemed to work, both because of the absurdity of the differences between the parties, but to the extent there were, I'm going to take "tastes great," that is, instant gratification, and David will take "less filling," and that is, longer term investment as our themes.

So I also want to take just a moment to reflect. I was last here about six months ago for an 80th birthday part for Robert Fogel of the University of Chicago, who was a great champion of the use of what is, certainly in my intellectual history, if not the intellectual history of the world, the first biomarker to be widely used in social, anthropological and other sciences, beginning in the early 19th century, where it clearly demonstrated big social disparities in health. Bob Fogel picked it up in an effort to understand access to nutritional resources, first in the context of slavery, quite controversially, to indicate that slaves actually had quite adequate nutrition, judging from their achieved heights, and then also to look at trends in broad historical time in what he thought would reflect changes in access to nutrition.

It's also been used by other economists such as Duncan Thomas to study children's outcomes, and in particular, look at how the allocation of resources within the family might affect some children differently, and whether mothers control resources or fathers control resources, how that affects kids' outcomes. Duncan and John Strauss have also used it to study effects on productivity in economies characterized by inadequate nutrition resources overall.

What's come out -- and I've done some of this work myself in the context of European history -- is the distinction that Bob Fogel does make in his work between nutrition and net nutrition. And in that model, food intake is only part of what achieves height. It also matters what are the demands on those nutritional resources, primarily from childhood disease, and diarrheal disease in particular.

Well, in looking at the broad sweep of a lot of European history, it's actually that side of things, public health, water quality, things like that, that matter a lot for child health. Access to food is often of secondary importance in determining those heights. But we would never have really focused on the importance of those public health innovations if we hadn't started from looking at this biomarker.

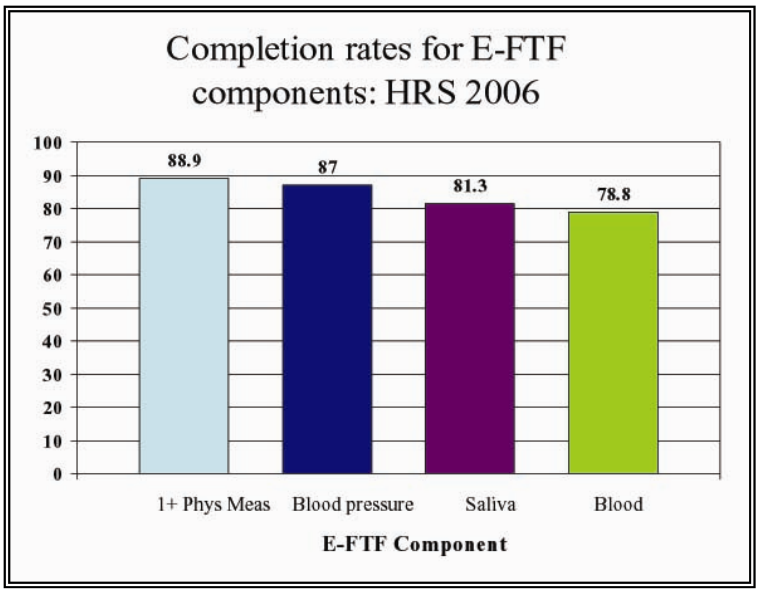
"So I think another thing that's going to happen, and NSHAP will add to this trend, is that there's a new item now on our research agenda. Now that we've finally put biomarkers on population studies, we're going to have to look at the data and compare them and try and understand why we find differences where we do find differences…"

So what have we learned from heights? A few things that I think we ought to bear in mind for other biomarkers. One is the importance of selection. A lot of the data sources used in the study of heights come from the military, which was obsessed with measuring people. In part that's because they valued tall people. That valuing took a number of different forms. In some places, like France, it took the form of a minimum height standard, so people below some cutoff never got in. That's pretty easy, if you know what that is, to make a correction for what you're missing.

In other contexts it was much more subtle, either actual

differentials in bonuses or payments for people signing up according to height or when military was actually a desirable, voluntary occupation, they could be choosy about who they took, and that could lead to differentials in observed height distributions as representations of the true population. So selection is an important issue, and I'll show you that in a minute in more modern context. Also, the biomarkers can have multiple interpretations. They can reflect economic differentials, differences in public health, and also, ultimately, genetic differences.

So now to HRS. HRS is, as many of you know, a cooperative agreement between the National Institute on Aging and the University of Michigan. It's a nationally representative longitudinal study. We're down now to about 18,000 people interviewed every two years in a population 50 and older. And until recently, the primary mode of interview was the telephone, which limits tremendously what you can do with biomarkers.



We did some early work with biomarkers in supplemental studies, one of which involved a rather expensive in-home evaluation of dementia that was called the Adams Aging Demographic and Memory Study. And in that study, of the 856 people that we went into the home to do the dementia workup, 850 gave us a DNA sample, so we had and enormously high cooperation rate once we got in to do this detailed interview. And that was used to determine APOE status, which is a known risk factor for dementia. But beyond that, it's just been stored and is available as a repository.

Secondly, we did a much less expensive supplement on diabetes that was a mail survey coupled with a mail request to do an A1c test by dry blood spot to be mailed in to the lab, and we got about 1,200 people to do those blood spots.

However, our big entry into the biomarker business was in this most recent wave of 2006, and the data from this, hopefully, will be out toward the end of this month or early July. What we felt was that to justify the extra cost and respondent burden of doing a face-to-face interview, as opposed to a telephone interview, we wanted to combine several related types of new information that, together, would offer a lot of new research opportunities.

So one key element are dry blood spot biomarkers, for which we've already done A1c, total cholesterol, HDL, and we will be getting C-reactive protein. We may also get some other things down the road, but those are the ones that we're hoping to complete soon. We also measured height, weight, and waist circumference. Height and weight we had asked people about, but we had never measured. Blood pressure we measured directly, and we also did some physical performance measures of grip strength, lung function, gait speed and balance.

A key element for broadening the research outreach of HRS was also adding a psychosocial questionnaire to get at things such as those that Noreen showed about social networks and social activities, but also, I think somewhat innovatively for a study of the scope of HRS, true measures of personality, the big five measures of control beliefs, measures of things like religiosity, perceived discrimination, a range of topics in social experience and psychological characteristics that I hope will turn out to be useful.

And then finally, we also collected, via saliva sample, DNA, which now resides, again, in a repository. This was done on half the sample, a random half, and we will be doing the other half, which, I guess, is no longer random in 2008. But by the end of 2008, everyone will have been approached to do this.

So what might we learn in the short-term? Well, clearly biomarkers can resolve some of the weakness of self-report data. They can correct inaccuracies in self-reporting, whether those are just lack of knowledge or recall or deliberate attempts to mislead. And secondly, diagnosis self-reports, even when they're as accurate as the person can possibly make them, hide two things,

typically. One is the effectiveness of treatment. So we ask people if they've been diagnosed with hypertension or diabetes, but we have no idea how well managed that condition is for that person compared to others. Secondly, of course, it tells us nothing about undiagnosed or possibly unreported disease.

What might we learn in terms of topics? Well, the obesity epidemic, which has become Public Health Enemy Number kind of 1b, after smoking, well, how serious is it, really, for health? Because most of the things that obesity produces are manageable through prescription drugs or other treatments.

So hypertension, blood sugar and so forth, a recent paper in *JAMA* showed that the risk factor load for an obese person, NHANES '99-2000, was about the same as for a lean person in the early NHANES of 1970. So medical management is kind of fighting against the obesity epidemic. Knowing how well that's playing out in the biomarkers of people, I think, will be extremely valuable. It will also, then, let us look at cohort changes in those cardiovascular risk factors.

Now clearly an important issue in any of these studies is the so-called SES gradient in health. This will help us understand how much of that is real and how much of that is reporting differences. Secondly, as Noreen also showed, it will allow us to examine what are the biologic pathways that link SES to health outcomes, or fail to do so, which will then direct our attention to look at other things. And finally, of course, down the road, better biomarkers will open up still new opportunities.

An example of what the biomarker can do I'm going to give from the diabetes study, a paper with Michelle Heisler in which we looked at racial differentials in A1c. Now race differentials in diabetes are enormous. They're enormous in prevalence, they're enormous in incidence, and they're enormous in staging of the disease. So the fraction of African-Americans on insulin is much higher than that of whites, for example.

And we know all that already in the HRS, but we don't know what their actual A1c scores are, and that measures what their average level of glucose is, how well the disease is being managed. And what we found was that there were, on top of the existing differentials in prevalence, incidence and severity, and while that explains a little bit of the differences in A1c, even controlling for all of that, there's persistent racial differentials in A1c, which also, somewhat to my surprise, was not explained by education. Education is related to glycemic control among people with diabetes, but race trumps it in the data. So that's, I think, an important finding we get from having this biomarker.

So a few results from 2006, really to convince you they're worth looking at and to kind of whet your appetite for analyses that may come. First we got, I think, pretty satisfying completion rates for these different elements of the enhanced interview. Close to 90% on the physical measures, 87% for blood pressure, and around 80% for saliva and dry blood spots.

Selection into Physical Measures: Grip Strength

Self-rated grip	% completed	Mean grip
Very strong	98.6	38.6
Somewhat strong	96.7	33.4
Somewhat weak	90.6	25.5
Very weak	76.0	21.0

HRS

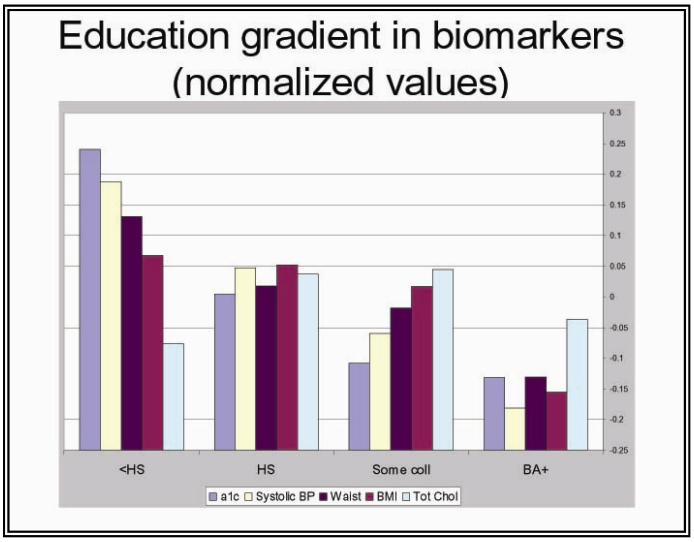
We did, however, observe, and we anticipated that we would observe this kind of selection into who gets measured. So grip strength, for example. People who have trouble with the muscles of their arms or have bad arthritis are going to be less likely to agree to do a grip strength test. That means, then, that the biomarker that we measure is measured on a selected part of the population. And this just shows you what we found in 2006.

We added a question that asked people, very simply, "How would you rate your hand strength?" And people who said their hand strength was very weak, and that's a small group, about four or five percent of the population, but they had much lower completion rates for the test, and those who did it had much lower grip strength. So with that we can do various kinds of re-weightings or imputations to do corrections for the fact that our measured grip strength represents, to some degree, a healthier part of the distribution.

Back to height, my first biomarker. And this was, to me, very interesting. What we found is that measuring

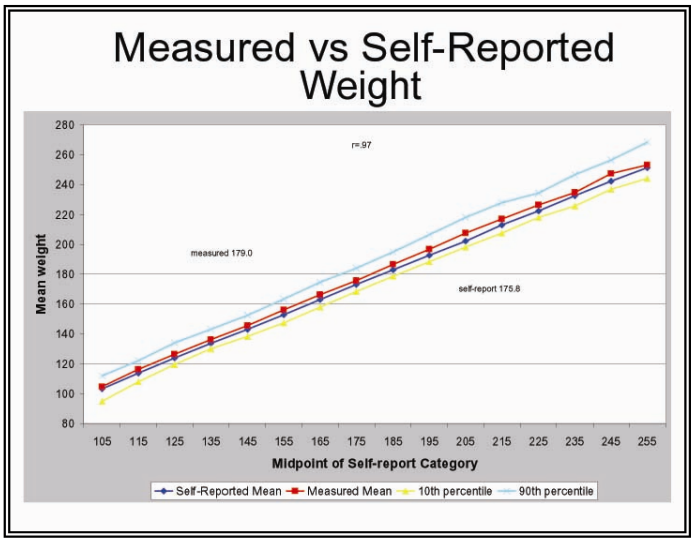
height in older populations is actually pretty important. And this is self-report in the same year as the measurement, and the correlation is just under .9. So not a super high correlation. And what you see here is the self-reported height is on the lower axis. The blue line is just the 45-degree line that represents exact match. So if the red line lay exactly on top of the blue line, we'd say measured heights and reported heights were exactly the same, but they're not.

Measured heights tend to be substantially lower. And the most obvious interpretation of this is that people remember their height when they were young, and that as they age, spinal compression and so forth, they shrink, and so they come out slightly shorter. Here's the



tenth and 90th percentile bands around that, and so you see, though, that the 90th percentile is still above the 45 degree line, so there are people whose measured height is higher than their self-report height. So some people, for whatever reason, are erring on the opposite side, reporting a height that's lower than their true height.

Another surprise was how well self-reported weight correlates with measured weight. So here the inaccuracy of self-report seems really pretty small. That may not be an argument for not doing actual measures of weight, but certainly I would argue that height is a much more important variable to measure than weight among the elderly. And again, the 90/10 percentile bands, even those at the 10th percentile, are below the mean of the measured, so again, there are people who are reporting weights below their actual rate. The errors are on both sides, even though, on average, there's a slight tendency to underreport weight.



Well, in response to Noreen, I was curious to see what kind of an SES gradient we had in biomarkers. If what you recall, what she said from Taiwan is that they found there was some correlation between the biomarkers and SES, but it was relatively mild. And what we find is something somewhat similar, so I normalized the values of all the biomarkers so I could put them on the same scale. So the scale is, essentially, fractions of the standard deviation in each of the measures.

And what you see is that for A1c, systolic blood pressure, waist circumference and BMI, there's a pretty steady socioeconomic gradient. Not that big. Less than high school education have about a .15 standard deviation above the mean, and the highest education, the lowest biomarker values, about .15 standard deviations below the mean. That's not an enormous spread, but it's in the expected direction. The one variable that does

not follow that pattern is total cholesterol, which is low for the very low educated, and then again for those with very high education. So that's going to take a little work to try and figure out what's going on with that.

So I think another thing that's going to happen, and NSHAP will add to this trend, is that there's a new item now on our research agenda. Now that we've finally put biomarkers on population studies, we're going to have to look at the data and compare them and try and understand why we find differences where we do find differences, and I think this will take us into looking at methods of measurement as well as issues of sample design and selection into measurement.

So here's a very preliminary comparison of our data from HRS 2006 with NHANES, pooling data from 1999 to 2004, the most recent data available. Notice, for one thing, the Ns. In general, one year of HRS, using half the sample, produces more observations on these biomarkers in the population 50 and older than six years of NHANES pooled. So even at two-year intervals with half the sample, will be a pretty powerful source for tracking trends in these biomarker values, assuming we can do them consistently, in a way that's comparable, from year to year.

Well, we come out extremely close on a number of these measures: on height, on waist circumference, on A1c scores. HDL, I think, is pretty close. Systolic blood pressure, slight difference, but really not disturbing. The ones in red are the ones that need a little bit of attention. Interestingly, our self-reported weights are actually closer to the NHANES measured weights than our measured weights. Because, if you remember, there was about a three or four pound difference, on average, between our measured and our self-reported weights. Our measured weights, it turns out, are almost four pounds heavier than NHANES.

Now that's, in some ways, not a really big difference, right? As the measurement of weight goes, it could be they were wearing pants and a sweater, you know, there's a lot of reasons for that. What's important to bear in mind, however, in our current obsession with obesity as defined by BMI of 30 or greater, the mean of BMI is now perilously close to 30 in the U.S. population, and it's a very normal distribution.

That means that the thick part of the distribution is right at the threshold. That means that small changes in your mean BMI push a lot of people over that threshold. So if you look at the percent obese, which isn't on here, it's different by six or seven percent of the population just from that small difference in measured weight. So one has to be pretty careful, and I'm not a big advocate of threshold measures. I think one needs to be a little more subtle in modeling than that.

"What's come out -- and I've done some of this work myself in the context of European history -- is the distinction that Bob Fogel does make in his work between nutrition and net nutrition. And in that model, food intake is only part of what achieves height. It also matters what are the demands on those nutritional resources, primarily from childhood disease, and diarrheal disease in particular."

Total cholesterol, we don't really know what's going on here, and I'm hopeful that we'll some other dry blood assay measures of cholesterol to do some comparisons to find out why we're substantially lower on total cholesterol than NHANES.

On diastolic blood pressure, people from NSHAP did share their data. Their data looked almost exactly like HRS. They used the same blood pressure machine as HRS. NHANES uses doctors and stethoscopes, and both our studies used machines. So it may be strictly a method of measurement difference that will have to be developed. There have been calibration studies that find nothing like a ten point difference in diastolic blood pressure, so it's a little surprising, but it's actually very nice to have two studies, so neither of us are out there on our own trying to figure out why we differ so much from NHANES.

So I think there's going to be a lot of interesting work just on this, and I think that's going to be important. I would argue economic studies of health do advance our understanding of health, that biomarkers will improve our measurement and modeling of health, including, especially, models of the interaction of health and economic variables such as those that Rena was talking about in her introduction. But I think that careful attention to measurement and sample selection will be needed to realize this potential. I think there's a risk of a lot of false findings if we're not pretty rigorous about looking at the data. Thank you all.

[*Applause.*]

aware of exactly the kind of horrible history that you're referring to, and do the best we can to be careful and mindful of all the problems and potential misinterpretations as we move forward.

Q: I should let the speaker have the last word, and this is unfair of me, so you can… Well, you mentioned cognitive intelligence. I mean, that's the prime example of how politics has corrupted findings about intelligence. And so it starts with very careful, sophisticated research people understanding some things and coming up with 0.25% of the variance due to that, but then that's public knowledge. And that public knowledge then becomes used by people with very considerable political agendas to advance a political…and if you think that's not going to happen with this research, then you're naïve, because it sure as hell has happened with intelligence.

And the results from intelligence testing is very stupidly used to look at population differences and describe population differences as then due to some inherited capacity for intelligence between minorities…it's always minority groups, of course, minority groups and so on. And any respectable researcher knows this is a bunch of nonsense, but yet it's promoted by all of the organizations that want to spend a great deal of money to promote their agenda. So…stay out of genetics. [*Laughter.*]

Closing Remarks

Lindau: We really should wrap up. Wow! That will make for an exciting last fifteen minutes of debate. I want to thank you all very much. I'm glad you stuck it out for the last session. When there's debate, that means we're interested. And that's, David, where you got the debate, right? [*Laughter.*]

Dr. Weir: You should have given him twenty minutes.

Lindau: The dialogue will continue. I'm sure there are plenty of economists and others who are venturing into the domain of combining biological with other disciplines who aren't here interfacing with the other disciplines doing this, and so to me, the more dialogue, the more opportunity for understanding consequences of being at this interface.

Raise your hand if you've completed and submitted your evaluation. [*Laughter.*] Okay, well, for those of you who have been honest and haven't yet submitted, you still have another chance. Thank you very much for attending.

Page Number	Speaker	Term	Meaning
1	Steptoe	ST-elevation myocardian infarction	when thrombus forms on a ruptured atheromatous plaque and occludes an epicardial coronary artery
1	Steptoe	angina	chest pain or discomfort that occurs when your heart muscle does not get enough blood
1	Steptoe	sclerosis	A morbid hardening of any tissue or structure
2	Steptoe	lumen	the hollow area within a tube
2	Steptoe	atheroma	a fatty deposit in the intima (inner lining) of an artery; can obstruct blood flow
2	Steptoe	thrombosis	a clot of coagulated blood attached at the site of its formation
5	Steptoe	fibrinogen	a protein present in blood plasma; converts to fibrin when blood clots
6	Steptoe	angiography	The use of X-ray images of blood vessels after injecting dye (contrast material) into the bloodstream. Used as a tool to diagnose many diseases affecting the arteries and veins, including inflammatory diseases.
7	Steptoe	power spectrum analysis	provides a quantitative noninvasive means of assessing the functioning of the short-term cardiovascular control systems
9	Steptoe	cannula	A hollow needle inserted into a blood vessel and used to give intravenous fluids
17	Lindau	NACDA	National Archive of Computerized Data on Aging
18	Lindau	transmucosal exudate specimen	Substance drawn from HIV antibodies out of the tissues of the cheek and gum
20	Lindau	dysplasia	Potentially precancerous abnormality of cervical cells
20	Lindau	NHANES	National Health and Nutrition Examination Survey

35	Luke	accelerometery	Used to measure tremor [usually originating in nervous system], in terms of amplitude (size) and frequency (speed) of the movements
35	Luke	calorimetry	measurement of quantities of heat
35	Luke	acti-cal	Brand of Physical Activity Monitors, which provide a quantifiable measure of Physical Activity and Energy (Caloric) Expenditure
53	Gudnason	MONICA survey	Multinational MONItoring of trends and determinants in CArdiovascular disease
54	Gudnason	Agatston scores	A cumulative score of the plaques appearing in a heart scan.
58	Lee	exons	sequence of a gene's DNA that transcribes into protein structures
58	Lee	SNP	Pronounced "snip," SNPs are single-nucleotide polymorphisms or one-letter variations in the DNA sequence.
74	Hendrie	statins	a class of drugs that help lower cholesterol levels in the blood.
87	Goldman	the Holter monitor	cardiovascular monitoring device