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 University of Chicago

and Northwestern University

Chicago - June 9 & 10, 2005

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The 3rd annual Chicago Conference on Biomarker Collection in Population-Based Health and Aging Research was organized and coordinated by Dr. Stacy Tessler Lindau, Dr. Jenna Mahay and Dr. Natalia Gavrilova, with assistance from Kathleen Parks, April Manalang, Adelle Hinojosa, and many others.

Financial support for the 2005 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. Additional support was kindly provided by the "Cells to Society": The IPR Center on Social Disparities and Health, Northwestern University.

The following proceedings were prepared from a transcript. Please bear in mind the informal nature of the workshop; participants' presentations and comments should not be used without permission.

ISBN 0-9769978-2-7

Suggested Citation

Lindau S, Gavrilova N (eds). Proceedings of the 2005 Chicago Workshop on Biomarkers in Population-Based Aging Research. NORC - University of Chicago Center on Demography and Economics of Aging. Chicago, IL. 2006.

WORKSHOP SPEAKERS

Emma Adam, PhD	Assistant Professor, School of Education and Social Policy, Human Development and Social Policy Program at the Northwestern University
John Cacioppo, PhD	Tiffany and Margaret Blake Distinguished Service Professor, University of Chicago
Chris Clark, MD	Director, Memory Disorders Clinic at the University of Pennsylvania Health Systems; Associate Professor of Neurology, Associate Director of the Alzheimer's Disease Center, University of Pennsylvania Medical School
Natalia Gavrilova, PhD	Research Associate at the Department of Obstetrics and Gynecology, University of Chicago and Chicago Core on Biomarkers in Population-Based Aging Research at the Center on Demography and Economics of Aging, NORC and the University of Chicago
James House, PhD	Professor of Sociology and Research Professor at the Survey Research Center Institute for Social Research, University of Michigan
Thomas Hummel, MD, PhD	Professor, Department of Otorhinolaryngology, "Riechen und Schmecken" (Smell and Taste Clinic) University of Dresden Medical School, Germany
Arun Karlamangla, MD	Assistant Professor, The David Geffen School of Medicine at UCLA; Physician in Internal Medicine and Geriatric Medicine, UCLA Hospital
Chris Kuzawa, Ph.D, MsPH	Assistant Professor, Department of Anthropology at the Northwestern University
Kenneth Langa, MD, PhD	Assistant Professor, Department of Internal Medicine at the University of Michigan
Stacy Tessler Lindau, MD, MAPP	Assistant Professor Obstetrics and Gynecology and Medicine-Geriatrics, University of Chicago; Director, Chicago Core on Biomarkers in Population-Based Aging Research at the Center on Demography and Economics of Aging, NORC and the University of Chicago
Jenna Mahay, PhD	NIA Postdoctoral Fellow, Center on Demography and Economics of Aging at NORC and the University of Chicago
Thomas McDade, PhD	Assistant Professor, Department of Anthropology, Northwestern University; Director, Laboratory for Human Biology Research

Douglas Richardson, BS	Director of Information Technology Group for the Department of OB/GYN at the University of Chicago
Narayan Sastry, PhD	Senior Social Scientist and Associate Director of the Labor and Population Program at RAND
Maxine Weinstein, PhD	Distinguished Professor, Center for Population and Health at Georgetown University
David Weir, PhD	Research Professor, Survey Research Center at the Institute for Social Research, University of Michigan

PREFACE

The 2005 Chicago Workshop on Biomarkers broadened our discussion from methods for biophysiological data collection in population-based research to theoretical foundations and analytical approaches to integration of these measures with self-report data. It also came to fruition via collaboration with Northwestern University, particularly Thomas McDade, and the newly initiated Cells to Society (C2S): Center on Social Disparities and Health.

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, and others) and researchers, many of whom receive funding from the NIA. 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. Finally, in an effort to draw innovation from outside academe, the workshop has also included an outside speaker, or "Translations," series. In 2004, an FBI special agent spoke on the collection of biological data from crime scenes, provoking thought and lively discussion about human subjects' rights and the importance of preventing "biologic contamination" of the scene, or research setting. Additionally, he shared biomarker collection technology and equipment used by the FBI, but unfamiliar to most researchers. This year, a NASA speaker, the first U.S. physician astronaut, spoke on not only the technical and technological challenges of collecting biological data from astronauts in space, but the sensitivity of such data; aberrations from normal could prove career-threatening. 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 NIA Chicago Core on Biomarkers in Population-Based Aging Research (CCBAR) in the 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 for participants, investigators new to this field and for alluring future conference speakers. We thank each of the individuals who shared and edited their presentation for publication in the *2005 Proceedings* and for the very engaged participants whose repeat attendance and intellectual involvement fuels both our scientific work and our enthusiasm for continuing this Workshop series.

Stacy Tessler Lindau, MD
Thomas McDade, PhD

INTRODUCTION

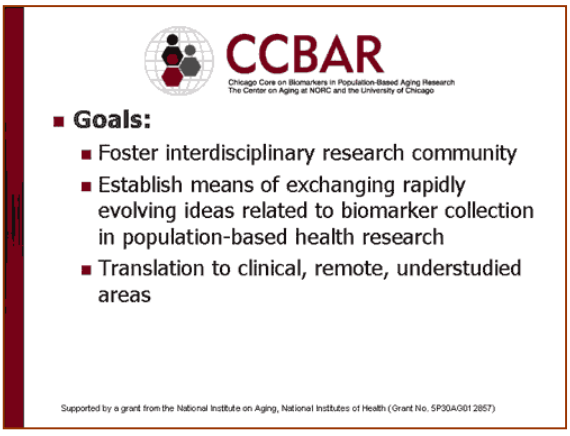
Collaborating to Maximize the Merging of Biomedical and Social Scientific Study of Population Health and Introduction to the Chicago Core on Biomarkers in Population-Based Aging Research

Stacy Lindau and Jenna Mahay

My name is Stacy Lindau. I know many of you, and some of you are new faces of people I've been communicating with over phones and emails for many months; and some are people I'm just meeting for the first time. So I want to thank you very much for coming this morning, and I think we have an exciting, excellent workshop, and I welcome you.

By way of acknowledgement, I want to give a brief introduction. This workshop was planned this year as a new joint venture between those of us at the University of Chicago who have been hosting the workshop for the last three years and, our friends at Northwestern University, particularly Thom McDade and his colleagues. This allows us to expand the focus and the collaboration for the workshop. Jenna Mahay at the University of Chicago worked very closely with me and with Thom McDade in organizing the meeting conceptually.

We've had many people who participated in planning the meeting, and they should not go without recognition. They worked very, very hard to make it all possible and for you to be here, and these include Arlene Dattels at Northwestern, Natalia Gavrilova, April Manalang, Kathleen Parks, and Nathan Sidles. Thank you to all of those people for making the logistics come together.



You all have a copy of the agenda in your folders and so that will tell you what you need to know about what will happen when. There is a change in the agenda for tomorrow morning. Unfortunately, Jack McArdle had a medical emergency and won't be able to attend. On the flip side, very fortunately, Ken Langa has been exceedingly gracious. He is going to fill that position by talking in part about the ADAMS study which is occurring in conjunction with the Health and Retirement Study.

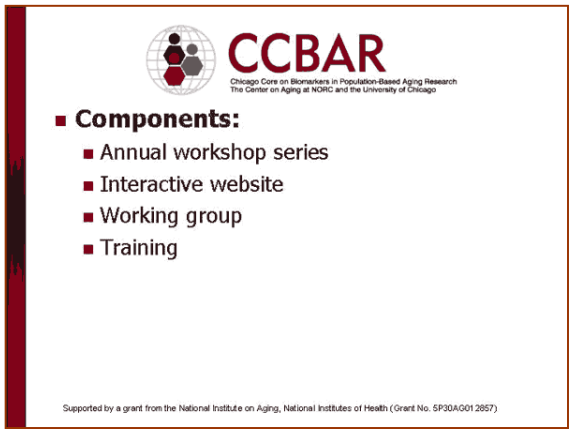
I want to briefly orient you to the Chicago Core on Population-Based Health Research. When we started this workshop three years ago, it was an outgrowth of the National Social Life, Health and Aging Project (NSHAP), which is a project funded primarily by the National Institute on Aging. We were needing help in thinking about how to execute the ambitious biomarker piece of that project and so we had a workshop inviting many people really to help serve as consultants to that project.

So some of you have that history in mind, and I want you to leave that history behind to some degree because really we're not here to talk just about the NSHAP project. Most of us are working on projects that integrate population-based social scientific research with biological and physiological data collection; and that's what brings us all together. Of course the perspective of NSHAP is in the room, as are the perspectives from many other studies; and so, we've grown.

Part of that growth has been to formalize the work of those of us who are interested in integrating biological and sociological markers, and that has occurred with funding from NIA through the support of the Center On Aging at the University of Chicago. The primary goal of this core is to foster an interdisciplinary research community interested in population-based health research, and we are accomplishing this through a variety of mechanisms, this workshop being one.

We're also aiming to establish a means for exchanging rapidly evolving ideas related to biomarker collection in population-based research. We know that this is a moving target and very much started with foundation laid by Teresa Seeman and the work she has been doing for NIA to catalogue biomarker collection methods and related information. We have continued to move from that foundation, trying to identify the biomarkers that people are working with now, identify those on the horizon, think about ones that might need to go into retirement, and to communicate those ideas. And this workshop will cover some of that ground.

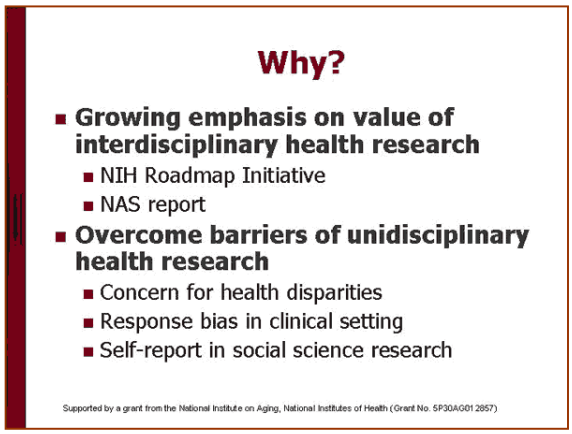
Then, finally, we're thinking about how to translate what we learn by developing methods for collecting biological and physiological data in the context of population-based research to the clinical setting. For example, do we need to do venipuncture to get lead tests on children, or could we be using finger stick methods to get lead tests on children? Certainly, we can imagine many ways in which clinical care of patients could improve by adopting some of the minimally invasive techniques we develop for use in the population-based research setting.



We are also thinking about translation of these technologies to accessing remote populations either for research or for diagnosis and treatment of diseases and also thinking about the applications to understudied populations, people who can't access health care, who don't, who are immobile, et cetera.

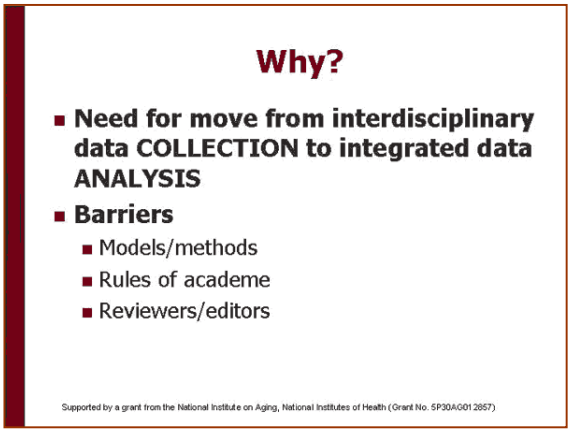
The components of the Chicago Biomarker Core follow: This annual workshop series has become a major part of our activity. We are developing an interactive website that we will tell you about later in the afternoon today. We host an interdisciplinary working group, held bimonthly at the University of Chicago. This attracts participants from our campus as well as Northwestern and beyond. Increasingly, the Biomarker Core is engaged in training, predominantly involving social science postdoctoral people interested in learning more of the biology and physiology side of health research. It will be interesting to see how that grows.

So why are we here talking about this issue? I see the crossroads very much being a community of social scientists who are interested in health research and a community of biomedical scientists who are interested in health research, and where they meet each other is in the population setting.



There's been, as you all know, growing emphasis on the value of interdisciplinary health research reflected in the NIH Roadmap Initiatives as well as a recent National Academy of Science report defining and describing interdisciplinary health research. Why should we do this? Well, in part, the way I think about it is it helps us overcome some of the barriers of unidisciplinary research. And when I say unidisciplinary, I mean biomedical alone or social science alone. Sometimes in my world in the medical setting multidisciplinary means a gynecologist talking to a pediatrician talking to an internist. In this case what I really am talking about are the large scientific disciplines. One motivation for this is the concern about health disparities; and as we know, there's been a lot of attention to funding of health disparities research.

The way I see this is that if we limit biomedical research to the clinical setting, we have a big response bias problem, and this is no big news to social scientists; but of course we end up studying disease and even risk factors for disease in populations who come to doctors. And this is obviously questionably generalizable to broader populations. So, for me, an important motivation to take the very costly and challenging approach for a clinician -- I have such an appreciation for what social scientists do through the NSHAP project -- but to do the work, to go out into the population setting, to learn from a generalizable sample, is a way to start to address some of the disparities in our understanding about health and our understanding about the onset of disease and maybe potentially how to treat disease in a population broader than just those people we see coming to the local hospital or clinic.

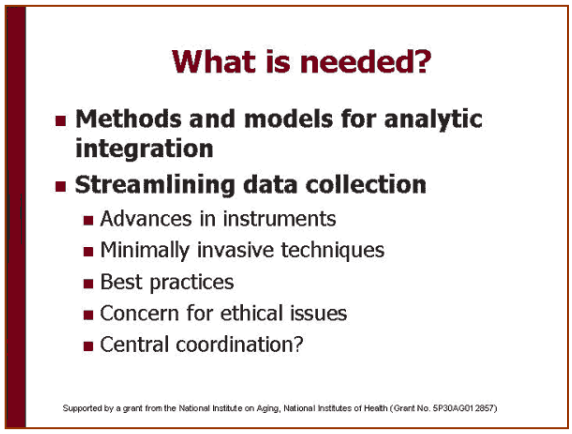


And then, of course, there's the limitation in social science research that's relied heavily on questionnaire-based self-report. So collecting objective measures of biology and physiology, even environment, physical measures of the environment, can help corroborate self-report, can help to synergize self-report to tell a different story. So these in my mind are the reasons why we're here today talking about these issues.

There is a need for a move from interdisciplinary data collection, which is what we've talked about the last few years, to integrated data analysis; and in our thinking about this working with Jenna Mahay and Erin York, for example, thinking about what are the barriers to getting from integrated data collection to integrated analysis and interpretation of findings. We now have several examples of large interdisciplinary groups of people studying health and collecting data about health; yet when we look to the literature, we don't find quite as many examples of truly integrated analysis of health problems. We think there are a number of reasons for this.

Some that we're looking into are the lack of conjoint models or methods for analysis or the rules of academe that require one to establish independence. I think this is changing but changing at different paces in different locales. Where integrated analysis is attempted, we may even face barriers with regard to peer and editorial review, limiting our options for publications.

Do the same reviewers have the capacity to critique an excellent demographic analysis and, simultaneously, the biological data that are presented? How do editors react to these things? How do scientific review committees react? So these are some of the barriers we need to think about; and while it's not formally part of one of the talks today, I think probably will be part of the discussion.



What is needed as we move forward? My sense is that we need methods and models for analytic integration of the sociological, demographic and biological data, and we need to continue to work on streamlining data collection. We all know about the costs involved with going into the population to collect data.

What advances do we need in instruments, both questionnaire-based instruments and biological and physiological instruments, advances in minimally-invasive techniques for collecting biological and physiological data and I think for collecting questionnaire-based data. I think we should apply the concept of minimally invasive to that as well.

I think we need to think about best practices. When should we collect biomarkers, when shouldn't we? Under what conditions should we? With what kind of training and background should we do this? I feel very strongly that we need to have a concern, an ongoing concern, and dialogue about ethical and human subject issues; and every time there's a new assay or new technique we have to revisit that issue. Every time we go to a new population we need to be considering the ethical side of what we do.

Finally, I suggest that we need to think about central coordination of biological data collection and storage. I put a question mark on the slide where it says "central coordination," but many of us are frustrated with the very decentralized way we have to work. We collect the data in a very decentralized manner (e.g. thousands of households across the nation) using a central infrastructure (i.e. a cadre of interviewers employed by a single survey research firm), but then we have to decentralize it to get the results back; should we move towards some mechanism for central coordination to maximize efficiency in the process?

These are some of the questions we are pondering and that will be covered here today. I'm going to wrap my talk up there, but I want to make some brief announcements.

Jenna Mahay

Thank you, Stacy. It's been a pleasure to be involved in planning this workshop, and it's been really interesting to watch the workshop evolve to meet the changing needs of the scientific community. Last year's workshop focused on the latest methods for collecting biomarkers from everything, finger stick blood spots to measures of sensory functions such as smell, taste and touch, to the best ways to measure cognition.

We included both technical and ethical discussions related to the collection of biomarkers in population-based research, and we look forward to continuing that dialogue, but this year we felt that there was also a need to go the next step and start thinking about what do we do with this information once we have it. How do we integrate biological data with other social measures in population-based research in order to truly understand health and aging from an interdisciplinary perspective? So it's great to see a lot of familiar faces and also new faces to contribute to this ongoing dialogue.

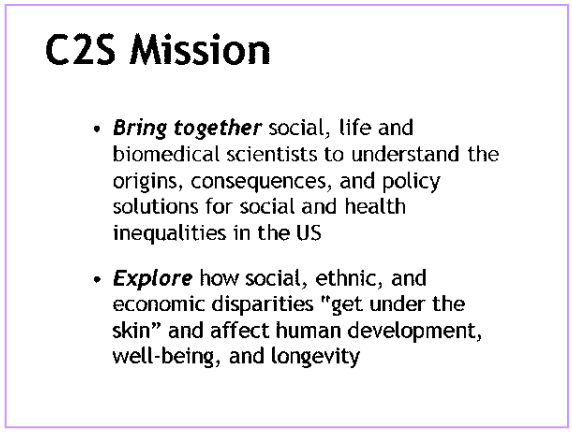
The feedback we got from last year also produced a change in the format of this year's workshop. This year we have fewer speakers but more time for discussion; and since this is a workshop we really hope that this will produce some insights that will help you with your own analyses and to really move the field forward.

We have also found collaboration among colleagues critical to the success of interdisciplinary research, and this year we're delighted to be collaborating with Thom McDade from Northwestern University in organizing this workshop. Thom McDade is an assistant professor of anthropology at Northwestern University and the director of the Laboratory of Human Biology Research. He's also been involved in helping to start a new center at Northwestern called Cells to Society, the Institute for Policy Research Center on Social Disparities and Health at Northwestern, and he's going to talk a little bit about that.

Description of the Center on Social Disparities and Health at Northwestern University

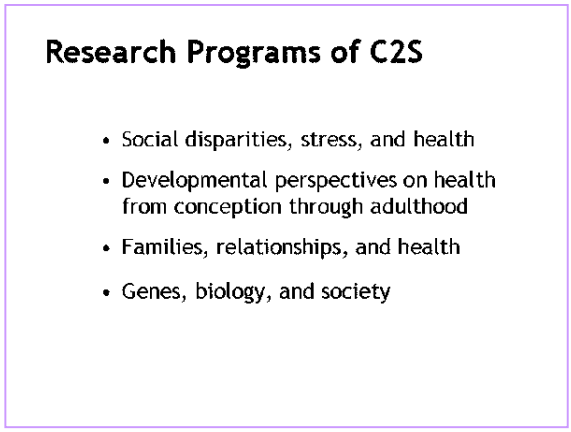
Thomas McDade

I'm grateful to Stacy and Jenna for extending an invitation to help co-organize this year's workshop. I've enjoyed participating the past couple years, and it's been fun this year to be part of the planning process.



I'm particularly excited by the agenda this year because the past couple years there have been a lot of technological developments, and we've done a pretty good job of selling people on the value and feasibility of biomarkers. Now we need to think critically about how to use these data and incorporate them in the service of a social science agenda, which is ultimately what many of us are interested in doing. What I'd like to do today is very briefly introduce you to a new center that we're starting at Northwestern, and then talk about a biomarker component of that center that may be of interest We would like to get some feedback from you about the directions we're considering taking the center in.

The center is called Cells to Society (C2S): Center on Social Disparities and Health. It's being coordinated by the Institute for Policy Research at Northwestern. Our mission is to bring together social, life and biomedical scientists around the problem of health disparities and the origins, consequences, and potential solutions for social and health disparities in the U.S. We're particularly interested in exploring how social, ethnic, and economic disparities get under the skin.

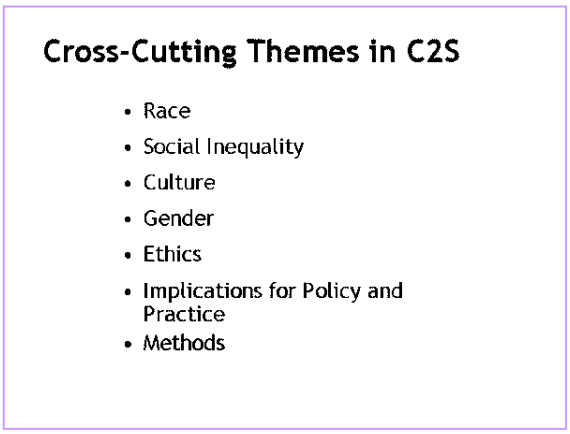


Health disparities is a common theme that links a lot of the people involved in the center and also reflects a broader interest by many of the participants in how biological processes are affected by social, cultural and economic environments.

Lindsay Chase-Lansdale is the director of the center. She's here with us today in the front row. She's in the program of Human Development and Social Policy at the School of Education and Social Policy at Northwestern. Emma Adam, who you will be hearing from later, is also a key member of the center, as is Chris Kuzawa sitting next to her. You'll hear from both of them later this afternoon; Tom Cook; Greg Duncan; Dorothy Roberts and Whitney Perkins Witt, who is right there, are also involved in the center. I know many of these names are familiar to many of you.

We've identified at this point four programs of research that draw on strengths that we have at Northwestern. One is stress and how social disparities affect stress with implications for health. A second program pursues

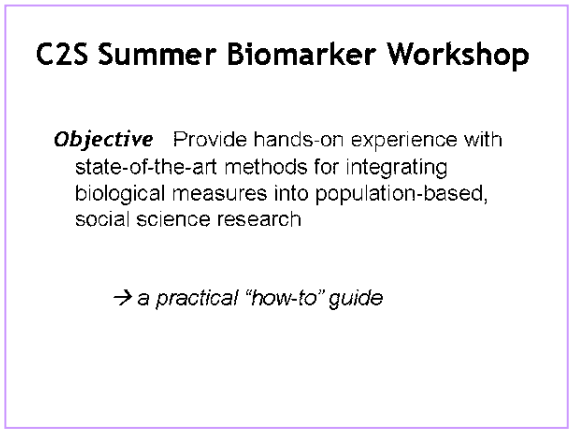
a developmental perspective on health from conception through adulthood. Chris is going to talk about that theme later today. Another is families, relationships and health; and Emma is going to talk about that also, with implications for stress. Lastly, genes, biology and society--or the broader political economic context in which our research takes place, and within which our society understands biology and health—is the final program of research.



There are a number of cross-cutting themes that go across these programs, including race, social inequality, culture, gender, attention to ethics and specific attention to implications for policy and practice which draws on the strengths of the Institute for Policy Research which has a history of major contributions in that regard. We are also devoting considerable effort to critical evaluation of methods, methods development and application for population level research.

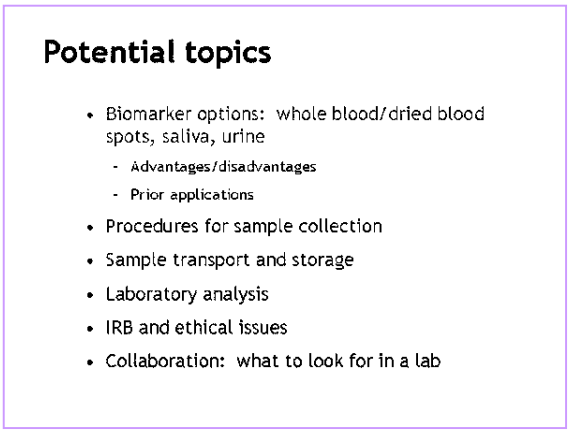
We launched the center on Monday. We had an introductory conference, and today is our sophomore effort as cosponsors of today and tomorrow's workshop, and next year we're launching a senior faculty research for senior scholars-- independent of discipline-- who are doing innovative research related to the theme of health disparities. So grab me quietly in a corner if you might be interested in talking about that. Otherwise look for the job announcement this fall.

This summer, we're intending to sponsor and organize a summer biomarker institute-- this is what I want to talk to you about, and also get input from you because a number of you and your colleagues are the key audience we're looking to serve with this activity. What we have in mind is something that we think is very complementary to what we've been doing here for the past three years, but with an emphasis on hands-on experience with state- of- the- art methods for integrating biological measures into population-based social science research.



We see this as more of a practical how-to approach, a week-long, lab-based format where you will learn how to collect a saliva sample, learn how to collect a blood spot sample, etc. We have a number of topics that we're considering, including: Discussion of the latest biomarker options, what you can measure, how you can measure it, and what body fluid you need to get to measure it; critical evaluation of the advantages and disadvantages of those measures and the options for collecting samples that are necessary to measure those things, and a discussion of prior applications of those methods; hands-on experience with sample collection; discuss issues related to sample transport and storage; and then laboratory analysis. We'll take our participants through the process of collecting that sample to transport, to laboratory analysis, to number, so people have a critical sense, a critical understanding, of what it takes to generate a concentration of CRP, or concentration of cortisol, so they know the limitations of the numbers that are given to them. Many social scientists are often shocked to realize just how much variation there can be. It's not as precise a science as some of us would like it to be all the time. And then of course attention to IRB and ethical issues, and discussion of collaboration and what to look for in labs that you may want to work with for the analyses that you may want to analyze.

We don't envision that participants in this workshop will come away as experts in biomarkers after a week with us. Instead, we're trying to educate people to be informed consumers so they'll feel more confident in seeking out collaborators and engaging in integrative research that brings biomarkers into their own personal research agenda.



That's what we have in mind, and what I'd like to hear from you informally as well as in your evaluation forms is this: Is this something that's of interest to you? Do you see a need, a niche for this? Because this is a lot of work frankly, and we're excited to do it, but only if there will be the demand and the interest in it. We, we want to make sure that's the case, and if so, what specific topics would you like to see featured there? What topics would be most useful for you and your colleagues, and what types of colleagues do you envision attending this kind of thing?

Would this be of interest to folks like yourself, or would it be more junior colleagues, graduate students, post-docs, people more junior in their career or more senior? We're curious what types of people you think this would be of interest to.

How do we coordinate this effort with what we're doing here today, the Chicago Biomarker Workshop? Should we try to put these things together? Is there enough overlap in the participants that we should coordinate these and make them back to back, or is eight, nine days way too much; and are these complementary and distinct enough that we should think of them as separate initiatives?

One other constraint is that the summer biomarker institute that we want to do will most likely have to happen in July because of constraints on our schedule; whereas this, the biomarker workshop, has typically happened in June. So one of those will have to move. We want to get a sense from you about what your feeling is about the relative timing of these things. I look forward to hearing from you about that.

Welcome to the workshop. We have a great agenda, and I'm really looking forward to today. Before we get started with our plenary speaker, we'd like to have all of you introduce yourselves.

INTEGRATING BIOLOGICAL AND SOCIAL SCIENCE DATA IN ANALYSIS

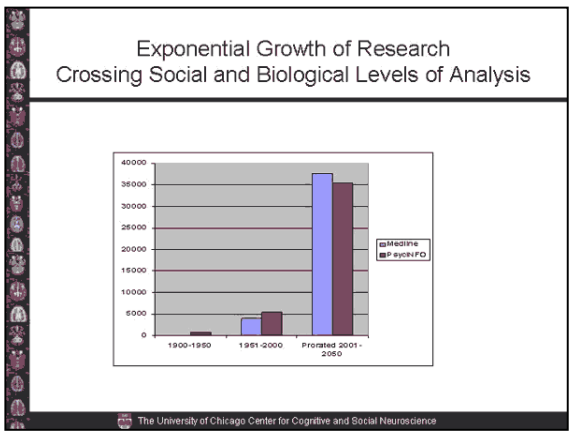
KEYNOTE ADDRESS: How do we Integrate Biological and Social Science Data in Analysis?

John Cacioppo

During the 20th century in the biological sciences, the architects of development and behavior were conceived as anatomical structures and genetic strings sculpted by the forces of evolution operating over millennia; the builders were cast as encapsulated within living cells far from the reach of social influences; and the brain was treated as a rational information processing machine. The additional information that might be attributable to the social world was conceived as best considered later, if the need arose.

Social factors, the reasoning often went, had minimal implications for basic development, structure, or processes of the brain or mind, in which case the consideration of social factors could be in fact entirely irrelevant. And even if relevant, the consideration of social factors was thought to render the study of the human mind and behavior too complicated to sustain scientific progress.

The embrace of the neurosciences by cognitive and social scientists throughout most of the 20th century was no less antagonistic. World wars, a great depression, and civil injustices made it amply clear that social and cultural forces were too important to address to await the full explication of cellular and molecular mechanisms. Biological constraints, mechanisms, and insights tended to be ignored, often under the misguided auspice of protecting the behavioral sciences from the onslaught of reductionism. Accordingly, antagonism between biological and social sciences characterized psychology throughout most of the 20th century.



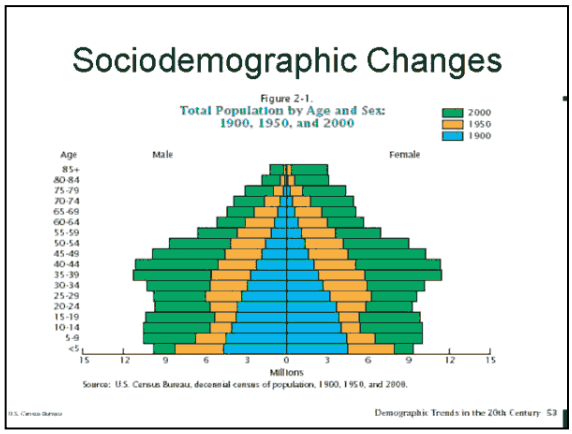
At the dawn of the 21st century, neuroscientists, cognitive scientists, and social scientists began to collaborate more systematically, bonded by the common view that the mind and behavior could best be understood by a multi-level analysis centered on the brain and biology.

Evidence for the historical trend I have described is easily obtained. For instance, in the left panel of the figure, I have depicted the number of articles in which the keywords "social" and "biology" or "biological" are found in Medline (designated in blue) and PsycInfo (designated in red) for the period 1900-1950. One such article was found for this period in Medline and 800 were found in PsycInfo. I repeated this analysis for the period 1951-2000, the results of which are depicted in the middle panel of the figure. Results revealed that in the second half of the 20th century there were approximately 4,000 articles in Medline and 5,400 in PsycInfo.

Finally, I performed the same searches for the period January 2001-May 2005. Since this period represents only a small fraction of the first half of the 21st century, I scaled the number of articles found in by Medline and PsycInfo for January 2001 – May 2005 to specify the estimated the number one would expect to appear in the first half of the 21st century. These estimates are presented in the right panel of the figure. These data clearly show the rise in interest and research crossing social and levels of analysis.

The growth in research crossing social and biological levels of analysis over the past few years in particular is testimony that the abyss between the neurosciences and social sciences can and must be bridged, that the mechanisms underlying mind and behavior will not be fully explicable by a biological or a social approach alone, and that a common scientific language grounded in the structure and function of the brain and biology can contribute to this endpoint.

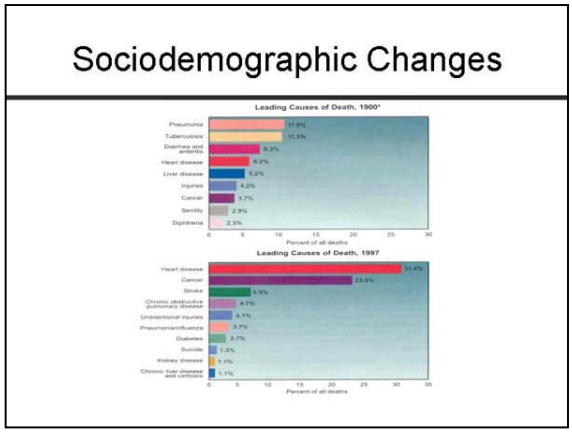
There are a number of factors fueling this perspective: the development of more smaller, more powerful computers; measurement developments (e.g., genetic, brain imaging, ambulatory, and blood-based measures); sufficient maturation of relevant scientific disciplines (e.g., consensus on overarching paradigms, nomenclature, and methodologies) to provide a solid base from which to launch interdisciplinary expeditions; the ready availability of mathematical tools (e.g., computational modeling, digital signal processing techniques, various multivariate statistics) and pressing societal problems that require an understanding of the interplay of social and biological processes.



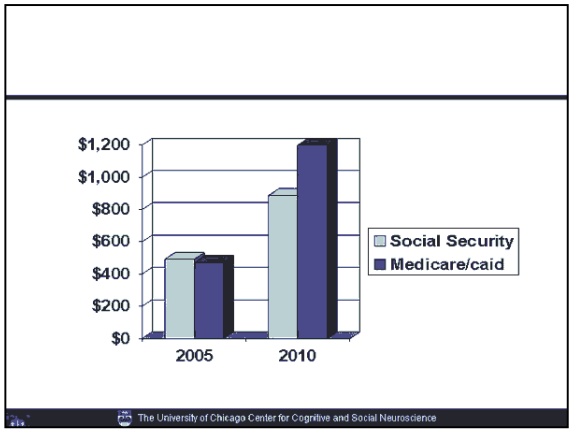
Over the last 100 years we have seen a dramatic change in the demography of industrialized nations. The figure (left) summarizes the population demographics in 1900, 1950, and 2000. These data clearly illustrate the aging of the United States (US) population. Moreover, the middle projections from the Census Bureau suggest that, by the year 2050, there will be more people over 65 years of age than under 14 years of age – an age distribution that is unique in human history.

Not surprising in light of these demographic changes, we also see a change in the causes of morbidity and mortality in the US. As depicted in the next figure, the causes of death in 1900 were largely germ-based disorders such as tuberculosis and pneumonia. By 2000, the major causes of mortality had changed from acute to chronic diseases such as cardiovascular disease and cancer. Furthermore, the changes in age, disease, and medicine have altered the economic burden of healthcare in the US. According to estimates from the Office of Management and Budget (2005), the annual federal costs of Medicare and Medicaid currently equal the entire annual costs of the Social Security program.

Within 5 years, however, the health costs are projected to greatly outdistance the rising costs of Social Security. Given the enormity of the projected economic burden, coupled with the now indisputable association between social and lifestyle factors and the development, progression, and maintenance of chronic diseases, it is essential that we gain a better understanding of the mechanisms underlying these associations so that better methods of prevention and treatment can be identified.



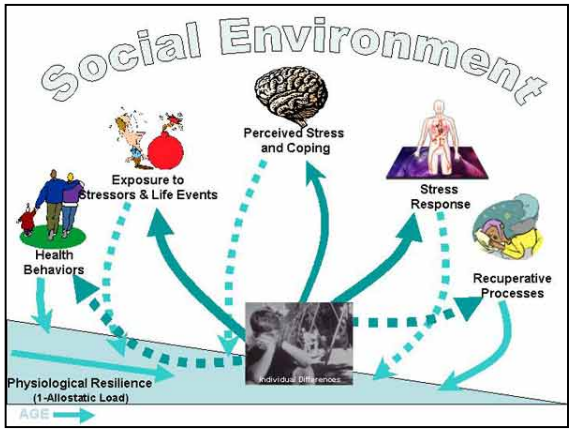
Adults tend to be less physiologically resilient and more socially isolated as they age. Research has shown that emotional closeness in relationships increases with age, as does positive affect until very late life. However, loneliness remains a significant complaint in middle-aged and older adults, and loneliness has been found to be a significant risk factor for depression in cross-sectional and longitudinal studies of older adults. Older adults therefore may be vulnerable to significant disruptions in social ties that can affect their feelings of loneliness and their mental and physical health. Causes of disruptions include death of a spouse or child, children leaving home or the area, giving up a home after the children have left, the loss of friends or neighbors through relocation or death, disability, spousal illness (e.g., dementia), divorce, reduced capacities for travel to see friends and relatives, dramatic cultural and technological changes, and bereavement from the loss of a child or spouse. Savikko, Routasalo, Tilvis, Strandberg, and Pitkala (in press) conducted a study of 6,786 Finnish adults 75 years of age or older and found that loneliness was associated with age, living alone or in a residential home, widowhood, low SES, poor health, poor functional status, poor vision or hearing, and rural rather than urban settings. Furthermore, they reported that the most common causes of loneliness were illness, death of a spouse, and lack of friends.



Sociodemographic and cultural changes in the U.S., including aging baby boomers, consistently high divorce rates, decreasing emphasis on the extended family, and gender differences in life expectancy contribute to the prevalence of loneliness. The increasing number of people living alone is changing the face of post-industrial societies. The average household size over the past two decades in the United States declined by about 10% to 2.5 (Hobbs & Stoops, 2002). By 1990, more than one in five family households with children under 18 was headed by a single parent, and within a single decade, the proportion of single parent households rose from 21% to 29% of all households in America (Hobbs & Stoops, 2002). Family households were not the only residential unit to become more socially isolated. There are also now more than 27 million people living alone in the United States, 36% of whom are over the age of 65 (Hobbs & Stoops, 2002). According to the middle projections by the Census Bureau (1996), the number of people living alone by 2010 will reach almost 29,000,000 – more than a 30% increase since 1980, with a disproportionate share of these being individuals over 65 years of age.

In our ongoing research, we have focused on objective social isolation, perceived social isolation (loneliness), and health. We found that lonely, relative to nonlonely, individuals are more likely to construe their world (including the behavior of others) as potentially punitive or aversive; they are more likely to be anxious, hold more negative expectations for their treatment by others, respond to quotidian events as hassles and stressors, appraise and cope differently with stressors, and adopt a prevention focus rather than a promotion focus in their social interactions. Together, these differences in social cognition predictably result in an increased likelihood of lonely individuals acting in self-protective and, paradoxically, self-defeating ways which, in turn, have been posited to activate social neurobehavioral mechanisms that may contribute to the association between isolation, loneliness, and mortality. We also found evidence for five transduction pathways. The resulting model – depicted below – begins with the established finding that physiological resilience – the

substrate upon which an individual draws upon to remain healthy –decreases with age. This is depicted in the figure below as the decline across time.



Pathway 1: Health behaviors. The transduction pathway that has received the most attention in prior research is health behaviors. Sometimes referred to as the social control hypothesis because the influence of significant others is thought to play an important role in moderating health behaviors, lonely individuals are thought to be exposed to weaker normative pressures from and control by friends and loved ones to perform healthy behaviors and to access health care when needed. Having fewer social ties also provides fewer sources of information, thereby decreasing the likelihood of having access to an appropriate information source to foster relevant health behaviors or to minimize stressful or risky situations. If lonely individuals are characterized by poorer health behaviors at least in part due to the influence of friends and loved ones who exert less pressure on them to adopt a healthy lifestyle, then differences in health behaviors may help explain the association between loneliness and health. This negative influence of loneliness on health behaviors is depicted in the model by the ascending dashed line (negative influence) between individual differences (loneliness) and health behaviors, whereas the positive impact of health behaviors on physiological regulation and resilience is depicted by the solid line (positive influence) between health behaviors and the physiological substrate. We found loneliness in middle aged and older adults to be associated with a few poorer health behaviors (e.g., higher percentage of their calories coming from sugar and fat) but the effects were relatively small, suggesting that other pathways are also important.

Pathway 2: Exposure to stressors and life events. The second pathway depicted in the model posits that lonely individuals are more likely to be exposed to stressors and major life events because they are less subject to social control (e.g., spouses and family who insist they eat well, diet, stop smoking) and, hence, these lonely individuals are more likely to exhibit poor lifestyles and accrue wear and tear on the physiological substrate. In an experience sampling study of young adults, which involved randomly beeping participants during a normal day to assess what they were doing and their perceptions and interpretations of their situations, we found that individuals who scored in the top, middle, and bottom quintile on the UCLA loneliness scale were exposed to the same objective stressors and circumstances during a normal day. As in our study of health behaviors, however, the lives of lonely young adults do not generalize to the lives of lonely older adults. For instance, in CHASRS we found that lonely older adults were exposed to more life events, including more frequent exposure to chronic and acute work and family stressors (Hawkley et al., in preparation). These influences are depicted in the model as an ascending solid line (positive influence) between loneliness and exposure to stress, and a dashed line between stress and physiological regulation and resilience (negative influence).

Pathway 3: Perceived stress and coping. These influences are depicted in the model as an ascending solid line (positive influence) between loneliness and perceived stress, and a dashed line between stress and physiological regulation and resilience (negative influence). The experience sampling study of young adults also indicated that lonely individuals perceived the hassles and stresses of everyday life to be more severe and the uplifts of everyday life to be less intense than nonlonely individuals. One possible explanation for this effect is stress buffering – that is, that individuals with dependable social ties are more likely to have access to others who can provide relevant assistance, support, comfort, or relief, and more likely to perceive support from the presence of others, when an individual is exposed to a life stressor. Differences in the participants' ratings of the severity of hassles and stressors remained, however, irrespective of the presence of others in the situation. In fact, social interactions, themselves a potential uplift and a source of pleasure for most

individuals, were not experienced as positively by lonely individuals. This result was found in our survey results, our experience sampling study, and in a recent fMRI study. In the latter, for instance, lonely, relative to nonlonely individuals, showed less activation of the caudate nucleus – a reward area in the brain – when pleasant pictures of people (e.g., a smiling couple, a skier on a mountain slope) were contrasted with equally pleasant and arousing pictures of objects (e.g., a panoramic sunset, a mountain cliff). These results point to differences in the emotional activation, cognitive appraisal, and coping in response to stressors.

Interestingly, nonlonely individuals have been found to do more active coping than lonely individuals, a coping strategy which on laboratory tasks has been associated with blood pressure responses governed primarily by increases in cardiac output (a "cardiodynamic" response served largely by beta-adrenergic and vagal mechanisms), whereas passive coping – the coping style that characterizes lonely individuals – has been associated with blood pressure responses characterized primarily by increases in total peripheral resistance (a "hemodynamic" response served largely by alpha-adrenergic and local vascular mechanisms). This research leads to a discussion of the next pathway, peripheral stress processes.

Pathway 4: Peripheral stress processes. These influences are depicted in the model as a solid line (positive effect) between loneliness and the stress processes and a dashed line (negative association) between these processes and physiological resilience. Social separation, rejection, or loss not only highlights an individual's lack of personal control but it places the individual at risk. We, therefore, posited that lonely individuals are not only unhappy, they have a heightened sensitivity to threats and attacks. Defensive behaviors such as rejection of others may help fend off treachery, rejections or attacks. Thus, although this self-protective focus may be paradoxically self-defeating in the long-term, it can serve to lessen the short-term damage of the negative social interactions lonely individuals are more likely to expect and to perceive in ambiguous interactions. Indeed, we found that lonely individuals showed smaller heart rate increases to a series of evaluative social stressors (e.g., speech stressors) than nonlonely individuals. Loneliness, however, was associated with chronically higher total peripheral resistance in both laboratory and ambulatory (Hawkley et al., 2003) settings in younger adults, and with age-related increases in systolic blood pressure in older adults.

Pathway 5: Recuperative processes: Detoxifying caustic days. Although the emphasis in prior research has primarily been on the toxic effects of stress and catabolic processes, we have posited that loneliness also diminishes the salubrity of restorative processes such as sleep. That is, we posited that loneliness may weaken the restorative effects of processes which serve to repair and maintain physiological functioning, foster recovery from stress, and contribute to the expansion of physiological capital and capacities as a function of adaptive transactions with the environment. Results from the prior grant period provide support for the notion that loneliness is associated with less efficient reparative and maintenance processes. We further posit that diminished sleep efficiency is associated with reduced physiological resilience and illness even when individual levels of stress are held constant. We have found evidence for this pathway in studies of younger adults and older adults. The pathway between loneliness and recuperative processes is designated in the model above by a dashed ascending line (negative influence), whereas the connection between recuperative processes and physiological resilience is depicted by a solid line (positive influence).

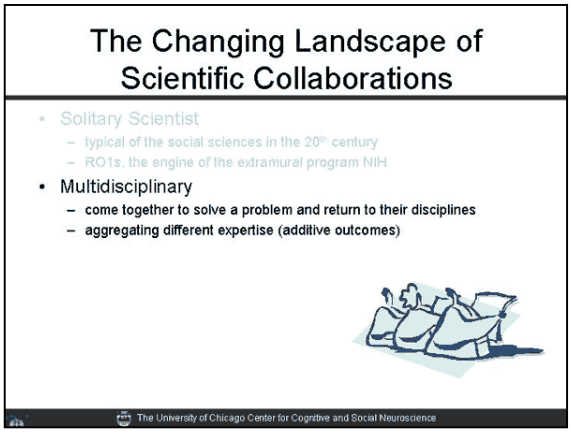
Although the past 15 years has seen a dramatic increase in the number of studies crossing multiple levels of organization, including social and biological levels of organization, it is useful to remember that such efforts are relatively new. In 1991 in an article in the American Psychologist on the future of psychology, Scott wrote:

"Psychology lacks a clear identity . . . Some of the vectors along which the subdisciplines have matured . . . have developed at obtuse angles to one another, and as the distance between them grows, they strain against the departmental membrane and are irritated by the requirements of common membership in a distended administrative unit. Social and biopsychology are an example. Most biopsychology students consider a core course in social psychology to be an impediment . . . I assume that our students in social psychology reflect that sentiment about their core experience in biopsychology." (T. R. Scott, 1991, American Psychologist)

A year later, Gary Berntson and I published an article in the same journal arguing for a very different future based the multiply and complexly determined nature of most social behaviors and the importance of understanding the underlying multivariate mechanisms.

Scott and we were in agreement that behavioral research was undergoing dramatic changes, though we differed in what we thought these changes might be. In broad brush strokes, it is useful to look at what are the changes in how investigators conduct their business and whether a moment of analytic thought about these changes might permit us to proceed in a more productive fashion.

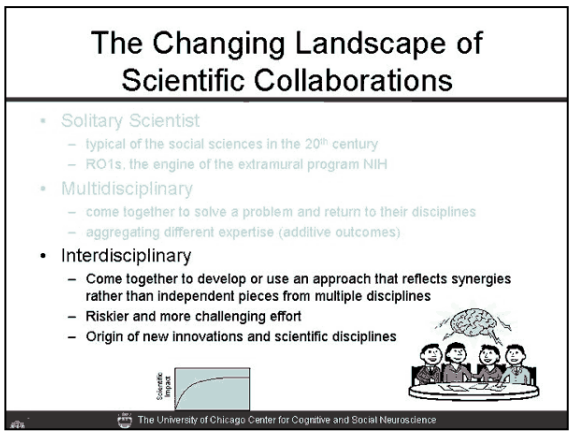
First, it wasn't long ago (as the quote by Scott suggests) that solitary scientists were the rule rather than the exception, and this tradition is still evident in many domains. The R01 mechanism, which forms the core of NIH extramural funding, was designed originally to support the solitary investigator. Important changes have been made in the R01 mechanism in recent years to make it easier for an R01 to support investigative teams at multiple institutions, but the R01 remains the mechanism of choice for solitary investigators and for most NIH institutes.



Complex questions began to exceed the expertise of individual investigators, and multidisciplinary research began to gain momentum. Multidisciplinary research is characterized by investigators coming together with their different expertise to solve problems then returning to their disciplines. That is, they aggregate their expertise, but their expertise is not changed or transformed, nor are the concepts or theories of their parent disciplines fundamentally changed.

Research on biomarkers could proceed along this route. This road to biomarker research would see experts in the social sciences calling upon experts in specific biological sciences to perform assays only then to treat the results of the assays as a simple metric of whether any difference in their social conditions was observed.

Interdisciplinary research differs from multidisciplinary research and is a better description of what is happening in the field of biomarkers. In interdisciplinary research, there are synergies rather than additive expertise across disciplines. It is decidedly riskier than multidisciplinary research because the latter requires only that one share the results of a common procedure to an investigator in another field. Interdisciplinary research requires greater innovation at the conceptual and operational level of research. If you're going to do something brand new and innovative, you're going to fail sometimes. If you never get to fail, then you're probably not asking interesting questions if you already know the answer. But when interdisciplinary teams succeed, they can solve what were thought to be intractable problems, influence parent disciplines, and produce new disciplines.



At the bottom of the slide (above right), I have graphed a negatively accelerating curve similar to the marginal utility functions found in economics. Along the abscissa is number of studies on a topic, and along the ordinate I've depicted the scientific impact of these studies. I would suggest that one reason interdisciplinary teams have greater potential impact is that they (typically) are breaking new ground in the questions being

asked. The more you ask the same kind of question, on average the less the potential the scientific value of doing so. Interdisciplinary teams may be characterized by impactful research in part because they tend to operate at a part of the rising rather than asymptotic part of this function.

Although the collaboration between Danny Kahneman and Amos Tversky was limited to two psychologists, Kahneman's characterization of their collaboration captures the synergism that can be achieved in the best of collaborations:

"Our principle was to discuss every disagreement until it had been resolved to our mutual satisfaction. Some of the greatest joys of our collaboration – and probably much of its success – came from our ability to elaborate on each other's nascent thoughts…" (Nobel Laureate Danny Kahneman, 2002, on his collaboration with Amos Tversky)

What might the ultimate product of research crossing social and biological levels of organization look like? E.O. Wilson in 1998 published a book called Consilience in which he suggested an answer to this question:

"The transition from purely phenomenological to fundamental theory in sociology must await a full, neuronal explanation of the human brain. Only when the machinery can be torn down on paper at the level of the cell and put together again will the properties of emotion and ethical judgment come clear . . . Cognition will be translated into circuitry. Learning and creativeness will be defined as the alteration of specific portions of the cognitive machinery regulated by input from the emotive centers. Having cannibalized psychology, the new neurobiology will yield an enduring set of first principles for sociology." (E. O. Wilson (1998). Consilience: The Unity of Knowledge)

Reductionism, a systematic approach to investigating the parts to better understand the whole, is sometimes confused with substitutionism – as depicted in E.O. Wilson's characterization of consilience. Reductionism, in fact, is one of various approaches to better science based on the value of data derived from distinct levels of analysis to constrain and inspire the interpretation of data derived from others levels of analysis. In reductionism, the whole is as important to study as are the parts, for only in examining the interplay across levels of analysis can the beauty of the design be ascertained.

Human biology is anchored in concrete anatomy and genetics, providing fundamental elements from which to draw interconnections and with which to construct theory. The social world, in contrast, is a complex set of abstractions representing the actions and influences of and the relationships among individuals, groups, societies, and cultures. The differences in levels of analysis have resulted in distinct histories, research traditions, and technical demands, leaving what some regarded as an impassable abyss between social and biological approaches.

All human behavior, at some level, is biological but this is not to say that biological reductionism yields a simple, singular, or satisfactory explanation for complex behaviors, or that molecular forms of representation provide the only or best level of analysis for understanding human behavior. Molar constructs such as those developed by social scientists provide a means of understanding highly complex activity without needing to specify each individual action of the simplest components, thereby providing an efficient means of describing the behavior of a complex system.

Chemists who work with the periodic table on a daily basis nevertheless use recipes rather than the periodic table to cook, not because food preparation cannot be reduced to chemical expressions but because it is not cognitively efficient to do so.

I am not the first to suggest reductionism should not be confused with substitutionism. Kuhlenbeck, a comparative neurologist in the middle of the 20th century, was certainly a reductionist but he also argued that:

Without the relevant unifying concepts, comparative neurology becomes no more than a trivial description of apparently unrelated miscellaneous and bewildering configurational varieties, loosely held together by a string of hazy "functional" notions. (Hartwig Kuhlenbeck, 1967)

Interestingly, if you look at the history of the term consilience, you find the term was first used in 1840 by Whewall in his book The Philosophy of Inductive Sciences.

"The consilience of Inductions takes place when an Induction, obtained from one class of facts, coincides with an Induction, obtained from a different class. This Consilience is a test of the truth of the Theory in which it occurs" (Whewell, 1840, p. 8)

Thus, Whewall defined consiclience as the jumping together of knowledge by linking of facts and fact-based theory across disciplines to create a common groundwork of explanation. Scientific knowledge, according to Whewell, is a consequence of human interpretation not the perception of absolute truths. Where Wilson's use of consilience is reductive, Whewell's is additive and synergistic. Where Wilson's points downward to more fundamental or true levels, Whewell points upward to higher generalities.

In 1992, Gary Berntson and I proposed what we call the doctrine of multi-level analysis, in which we argued that comprehensive understandings of many complex social behaviors would benefit from multiple levels of data being integrated simultaneously. We further articulated three principles as part of that doctrine: The principle of multiple determinism, the principle of additive determinism and the principle of reciprocal determinism.

The principle of multiple determinism specifies that a target event at one level of organization, but especially at molar (e.g., social) levels of organization, can have multiple antecedents within or across levels of organization. On the biological level, for instance, researchers identified the contribution of individual differences in the endogenous opioid receptor system in drug use while on the social level investigators have noted the important role of social context. Both operate, and our understanding of drug abuse is incomplete if either perspective is excluded. Similarly, immune functions were once considered to reflect specific and nonspecific physiological responses to pathogens or tissue damage. It is now clear that immune responses are heavily influenced by central nervous processes that are affected by social interactions and processes. For instance, the effects of social context now appear to be powerful determinants of the expression of immune reactions. It is clear that an understanding of immunocompetence will be inadequate in the absence of considerations of psychosocial factors. Major advances in the neurosciences can derive from increasing the scope of the analysis to include the contributions of social factors and processes.

A corollary to this principle is that the mapping between elements across levels of organization becomes more complex (e.g., many-to-many) as the number of intervening levels of organization increases. For instance, if there is a many to many mapping from the biological to behavioral and behavioral to the social levels, then investigating each of these mappings is a complex undertaking. If one starts skipping levels of analysis (social to biological), then the complexity of the problem increases geometrically. That is, the likelihood of complex and potentially obscure mappings increases as one skips levels of organizations. Accordingly, interdisciplinary teams benefit from having all the appropriate levels of expertise needed to undertake the mapping.

The principle of nonadditive determinism specifies that properties of the whole are not always readily predictable from the properties of the parts. Consider an illustrative study by Haber and Barchas (1983), who investigated the effects of amphetamine on primate behavior. The behavior of nonhuman primates were examined following the administration of amphetamine or placebo. No clear pattern emerged between the drug and placebo conditions until each primate's position in the social hierarchy was considered. When this social factor was taken into account, amphetamine was found to increase dominant behavior in primates high in the social hierarchy and to increase submissive behavior in primates low in the social hierarchy. The

importance of this study derives from its demonstration of how the effects of physiological changes on social behavior can appear unreliable until the analysis is extended across levels of organization. A strictly physiological (or social) analysis, regardless of the sophistication of the measurement technology, may not have revealed the orderly relationship that existed.

The principle of reciprocal determinism specifies that there can be mutual influences between microscopic (e.g., biological) and macroscopic (e.g., social) factors in determining behavior. For example, not only has the level of testosterone in nonhuman male primates been shown to promote sexual behavior, the availability of receptive females influences the level of testosterone in nonhuman primates. Accordingly, comprehensive accounts of these behaviors cannot be achieved if the biological or the social level of organization is considered unnecessary or irrelevant.

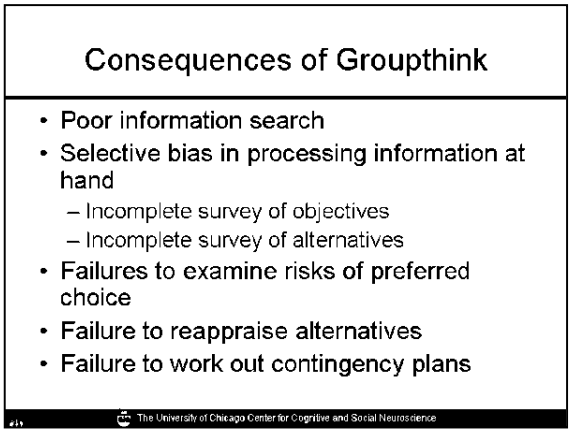
I've thus far tried to convince you interdisciplinary teams are an important wave of the future, but let me now turn to why interdisciplinary teams of scientists are not always better. By considering those cases where interdisciplinary teams can fail, my hope is that we will be in a better position to design interdisciplinary teams that overcome the inherent problems in group processes.

One frailty in "social" cognition – that is the cognitive operations that are produced from an interacting set of interdisciplinary scientists – is known in social psychology as the Hiring by Committee Dilemma. Briefly, imagine that there are six of us from different departments who have come together to hire one of two candidates, Sally or Sam. We all know three positive facts about Sally and three negative facts about Sam. That is, the common knowledge favors Sally by a score of +3 for Sally to -3 for Sam – an easy choice on the surface. Of course, since this is the shared or common knowledge, there was no need for our special expertise on the topic.

But there is more to this decision task, just as there is in interdisciplinary teams. We each have unique knowledge, as well. In this committee task, imagine that we each have two pieces of unique knowledge: each of us knows one negative fact about Sally and one positive fact about Sam. That is, the unique knowledge favors Sam by a score of +6 for Sam to -6 for Sally.

As should be apparent, if all of the information is presented and discussed, Sam wins with a net score of +3 for Sam and -3 for Sally. In fact, what studies using this paradigm show is that Sally rather than Sam tends to be hired. Just because different knowledge bases are brought together into the same research team does not mean that all the relevant knowledge is presented or used in the group's problem solving. Common knowledge can often carry more weight, for reasons you can imagine. You're much less likely to suggest a dissenting view if everyone already agrees, especially when even when the information you have favors Sally over Sam by a 3:1 margin. People censor themselves for informational and normative reasons.

This example illustrates that group problem solving, such as that which characterizes interdisciplinary scientific teams, is not necessarily better. Common knowledge tends to be presented first and tends to outweigh unique knowledge even though, ironically, the reason for having an interdisciplinary team is the unique knowledge. The implication is that these interdisciplinary teams can actually be quite ineffective if they do not try to ensure everyone's view is explored, all of the information is brought out, and people feel comfortable raising nonconsensual issues.



Second, interdisciplinary scientific teams can be worse than individuals because of a process called "groupthink." Examples of groupthink abound across US history: the Bay of Pigs Fiasco, the Challenger disaster, the Vietnam war, invasion of Iraq in 2003.

Among the consequences of groupthink (see Slide 23) are poor information search, selective bias in processing information at hand because of an incomplete survey of objectives and an incomplete survey of alternatives, failures to examine risks of preferred choice, failure to reappraise alternatives, and failure to work out contingency plans.

Among the processes underlying groupthink (see Slide 24) are an illusion of invulnerability (e.g., due to high intelligence), self-appointed mind-guards, direct pressure on dissenters, self-censorship, illusion of unanimity, collective rationalization, stereotyping of outgroups (those who disagree with or criticize you), and a belief in the inherent morality of the group.

Interdisciplinary scientific teams are not immune to the factors that contribute to groupthink: high cohesiveness, insulation of the group, strong leader who does not explicitly value dissenting perspectives (even though the leader may value them implicitly), and high stress (e.g., time pressure) with a low degree of hope for finding a better solution than the one favored by the leader or other influential person.

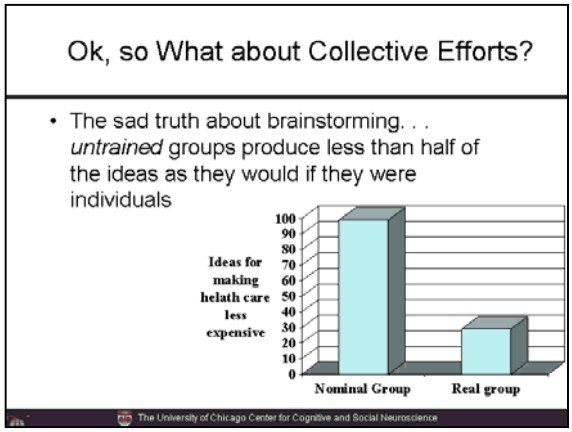
If groupthink is a feature that we have to be concerned about, what might you do about it? There are several things. One is to not be wed to a particular outcome, solution, hypothesis, or interpretation. The joy of disconfirming evidence is that you get to learn something. If your hypothesis is confirmed, you learn less in may cases than if your hypothesis is incorrect.

Another feature of interdisciplinary teams that lessens the likelihood of groupthink is a weak hierarchy. The leader is a learner. If the leader is all-knowing, then there are pressures for self-censorship or risk rejection by the valued group. If the leader is also a learner and is perceived by the group to be a learner, then this pressure is lessened.

Critical evaluation also needs to be valued in interdisciplinary teams if groupthink is to be avoided. Confirmatory and disconfirmatory data are treated equally. I worried early in my career when I saw myself building on data that supported a hypothesis but going back to make sure there were no coding or transcription errors in the data when it disconfirmed the hypothesis. I realized that doing this could potentially introduce a systematic bias, so I quickly learned to take the time to verify all the data and processing steps prior to testing the hypotheses, and to equally embrace confirmatory and disconfirmatory evidence.

A fourth feature that diminishes groupthink is the open allowance and consideration of minority opinions and views. Many of you know that one's worst critic in academics can be invaluable because he or she is the person most likely to note the problems in one's thinking, the alternative interpretations that were inadvertently overlooked, or relevant work in the literature that was missed. We might want to curse them,

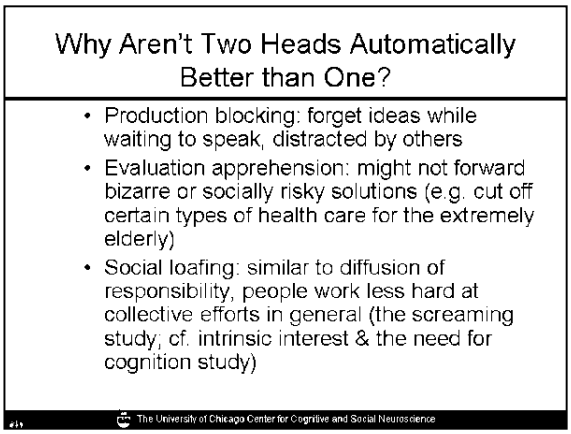
but in fact we couldn't do nearly as well without them. The same is true for critiques within interdisciplinary teams.



Finally, time pressure promotes groupthink and errors in collective problem solving, yet we oftentimes find ourselves in situations (e.g., grant deadlines) in which time pressures are sorely felt. To the extent we can design interdisciplinary teams that are not working under constant time pressure, one will have a better chance to be playful with ideas, to consider alternatives, to hear things that aren't perhaps going to move us to the very next step but in fact may keep us from making a step in the completely wrong direction.

What about brainstorming? The sad truth is brainstorming is also diminished in groups compared to individual action. Again we have evidence that bringing together an interdisciplinary team of brilliant experts will not necessarily lead to a better outcome. As above, considering why brainstorming can be less effective in groups than in individuals may allow us to construct working environments for the groups that allow one to optimize the considerable potential interdisciplinary teams offer.

There are a number of reasons groups on average perform more poorly than individuals on brainstorming tasks. One reason is production blocking. If a group is listening to one person's creative suggestion, the others in the group spend their cognitive efforts listening to (and being biased by) that person's suggestion rather than spending this cognitive effort on their own brain storming effort. In fact, they might also forget aspects of what they had been thinking, reducing the number of independent lines of creative thought that will be generated.



A second reason groups brainstorm less effectively than individuals I evaluation apprehension. Evaluation apprehension in brainstorming settings can cause people to edit what they think others would consider bizarre or risky solutions even though perhaps those are the very options about which the group most needs to hear.

Third, this is social loafing; people can freeload or exert less cognitive effort when brainstorming because others are also responsible for generating ideas. In an early study of social loafing by Bibb Latane and colleagues, students sitting in a large room, each with a microphone in front of them and each wearing headphones that made it impossible to hear anything other than the experimenter's instructions. All students were told to scream into their microphone as loudly as they could. In half the conditions, the students were also told that each microphone was individually calibrated so that the Experimenter could measure how loudly each individual was yelling into his or her microphone. In the other conditions, the students were told that the microphones were calibrated to quantify how loud the group yells. The results showed clearly that the students yelled much more loudly when it was their own production rather than the group product that (they thought) was being measured. Social loafing can occur in interdisciplinary scientific teams, as well.

What can be done to minimize the negative effects of group dynamics on brainstorming? One is to encourage individual brainstorming followed by group brainstorming. This minimizes production blocking and, in the right contexts, evaluation apprehension.

We have discussed the need for and potential of interdisciplinary scientific teams, and we have considered three major obstacles to optimal interdisciplinary scientific research: the weight given to consensual over individual knowledge, groupthink, and the potential inhibition of brainstorming/innovation. Another set of more obvious but no less important obstacles also exists. First, participants in an interdisciplinary team need to achieve a common language. If a sociologist is speaking for the first time about stress to an immunologist, it may take you awhile for the two to communicate effectively because the associations each has to the term "stress" might have little overlap. Aspects of our communication with colleagues that we take for granted, such as the historical reasons for our paradigms and the implicit assumptions underlying our methods or analytic procedures may be opaque to a scientist in another discipline. Second, the epistemologies themselves may be quite different. How each scientist thinks about the scientific enterprise may need to be bridged. The epidemiologist may deal in associations and covariations, whereas the experimental psychologist may only study what can be experimentally manipulated. Neither is better than the other, each is best to address a certain kind of question.

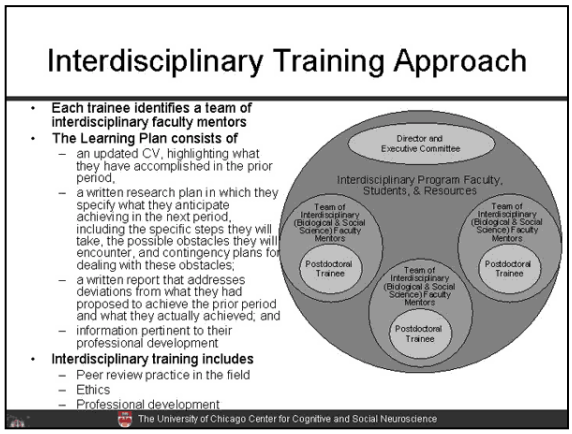
Both differences in language and differences in epistemology can be strengths rather than weaknesses if the investigators have the time and opportunity to work through the features that make effective communication and inquiry difficult at first. Forcing oneself to articulate implicit assumptions, for instance, or to explain oneself in nontechnical language, can be informative for oneself as well as helpful to one's interdisciplinary colleagues.

Third, interdisciplinary research is more effective when the participants adopt a learning posture rather than a defensive posture (or "show them how critical and therefore how smart I am" posture). Various observers of interdisciplinary study panels at NIH have observed that people tend to defend the value of their own disciplinary perspective and to be critical of others. People do this within their own disciplines, as well, especially when they are confronted by a divergent theoretical perspective.

The problem is not critical thinking but defensiveness. Above, I emphasized the importance of open and critical thinking and communication in interdisciplinary teams. But such a mode of thinking can also embrace alternative views because no one is wed to the positions and ideas that are being considered. Defensiveness, on the other hand, is caustic in interdisciplinary scientific groups because it promotes many of the problems in group dynamics outlined above.

And chemistry matters. You can have good people who don't fit together. It may be better to change the constitution of the group, not because anybody is doing anything wrong, but because the chemistry of the assemblage is problematic.

Training models are also undergoing changes as interdisciplinary research that crosses levels of organization becomes more common. I do not believe a single training model is best in all situations or for all individuals. For instance, does one train in an interdisciplinary field or in multiple disciplinary fields? Generally, in my opinion, the answer to this question is that it depends on what the status of the interdisciplinary field. If the field new and still emerging, trainees may benefit from having a solid disciplinary foundation; if the interdisciplinary field is well developed, then study centered in that interdisciplinary field carries less risk for the trainee. Whatever the training model, it is important to recognize that interdisciplinary training is a continuing affair, that additional disciplinary and interdisciplinary expertise requires continued investments in terms of time and effort.

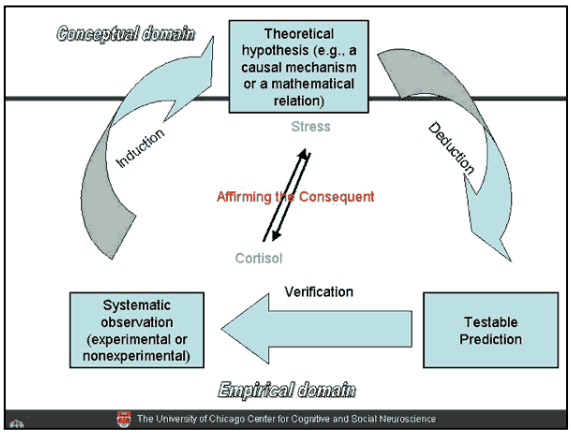


Slide (left) depicts one possible interdisciplinary training model. This interdisciplinary team approach to program direction is represented at the level of the individual postdoctoral trainee, as well. The trainees may be predoctoral or postdoctoral trainees, and the mentoring team, drawn from different apposite disciplines, can be customized to fit the circumstance and individual. Specifically, each trainee identifies a team of interdisciplinary faculty mentors (who will also be the trainee's primary research mentors). The training team, for instance, may differ for each trainee but cohere within a program of trainees due to an overarching administrative structure, quarterly meetings with the executive committee of the training program, and annual faculty meetings with all of the team leaders. In addition to the training activities, each trainee develops a "Learning Plan," which s/he reviews quarterly with their team of interdisciplinary faculty mentors and the Executive Committee. The Learning Plan consists of (a) an updated CV, highlighting what they have added (e.g., published) in the prior quarter, (b) a written research plan in which they specify what they anticipate achieving in the next quarter, including the specific steps they will take, the possible obstacles they will encounter, and contingency plans for dealing with these obstacles; (c) a written report that addresses deviations from what they had proposed to achieve the prior quarter and what they actually achieved; and (d) information pertinent to their professional development. These meetings not only send a clear message about expectations for professional development but they serve to identify potential problems at the earliest possible stage and devise solutions to them as well as to promote the development of realistic research objectives and professional growth.

Finally, given numerous demands, it is understandable why scientists might adopt measures or methods without thinking exhaustively about the underlying assumptions of these operationalizations or how they fit within a broader set. Such constraints are the stuff of which apocryphal stories are fashioned by philosophers of science. More than 60 years ago, Sir Author Eddington (1939), in his book A philosophy of the physical sciences, told the story of a hypothetical scientist who sought to study the fish in the seas. The scientist wove a two inch mesh net and commissioned a ship on which to sail the seas. Once on the high seas, this individual sailed to various sites, lowered the nets, hauled in a catch, measured and cataloged each fish, returned the catch to the deep, folded the nets, and sailed to another site to repeat the procedure. After several years of investigation, the scientist returned to announce there were no fish smaller than 2 inches in the seas. Taking the time to discuss the measures and methods we use, the underlying assumptions, their strengths and limitations, and the theoretical implications of these operationalizations may pay dividends in the long term in science.

For instance, interdisciplinary research that crosses biological and social levels of analysis raise issues about how one might productively think about concepts, hypotheses, theories, and theoretical conflicts, and theoretical tests across levels of analysis. As John Platt (1964) noted in his Science article entitled "Strong inference":

We praise the "lifetime of study," but in dozens of cases, in every field, what was needed was not a lifetime but rather a few short months or weeks of analytical inductive inference … We speak piously of taking measurements and making small studies that will "add another brick to the temple of science." Most such bricks just lie around the brickyard. (John Platt, 1964, Science, p. 351)



Platt emphasized the importance of hypothetical deductive logic in scientific research. But is this a sufficient solution for nascent interdisciplinary scientific fields or fields with a central objective of diagnostics.

For instance, the field of biomarkers has as a major objective to infer biological, behavioral, psychological, or social states or conditions based an economical biological measure. Hypothetico-deductive logic is characterized by the articulation of two competing theoretical hypotheses which make contrasting empirical predictions, the collection of the relevant empirical evidence, and the rejection of the hypotheses whose prediction was incongruent with the observed empirical evidence. Research in fields such as biomarkers, on the other hand, emphasize inductive processes: A biological measure (e.g., salivary cortisol level) known to correlate with diagnostic category (e.g., a hypothesized state or condition such as "stress") is interpreted as evidence for that state when empirically observed. This form of inference is problematic, however. Even if we knew variations in stress were associated with corresponding variations in salivary cortisol, inferring stress based on cortisol represents a logical error called affirming the consequent because it denies the possibility that there are other antecedent conditions that could also produce variations in cortisol.

How might one deal with this situation? S. S. Stevens (1951), the editor of a Handbook of Experimental Psychology that influenced at least a generation of psychologists, suggested that when a correlate such as stress and cortisol was found, the next step was obvious:

The scientist is usually looking for invariance whether he knows it or not. Whenever he discovers a functional relation between two variables his next question follows naturally: under what conditions does it hold? The quest for invariant relations is essentially the aspiration toward generality, and in psychology, as in physics, the principles that have wide application are those we prize. (S. S. Stevens, 1951, p. 20)

I suggested earlier in this talk that social states and outcomes tend to be multiply rather determined and are subject to contextual (e.g., cultural) influences. To the extent that this is the case, the notion that one is dealing with an invariant relationship can be counterproductive. The diagnostician who finds stress and cortisol are correlated and then searches for the conditions under which it holds will find little order in the data because so many other factors (e.g., time of day, time since consuming food) will be unrecognized and uncontrolled. On the other hand, if the diagnostician who finds stress and cortisol are correlated next asks what is the specificity (rather than the generality) of this association, other antecedent conditions that influence cortisol are more quickly recognized, and contexts in which these other antecedents are controlled or quantified can be developed to allow strong inductive inferences about the state of an individual's stress based on cortisol levels. That is, the sensitivity and specificity of the biomarkers may be context dependent, and attention to these issues improves the quality of inductive inferences.

The products of these inductive inferences are abstractions that resides in the conceptual domain and serve as the ingredients of scientific theory. The scientific method relies on both deduction and induction, each in turn, to develop comprehensive scientific theories of a given phenomenon. I would like to end today by speaking briefly about two broach philosophical approaches to the abstractions and theories in the conceptual domain: scientific realism and scientific instrumentalism.

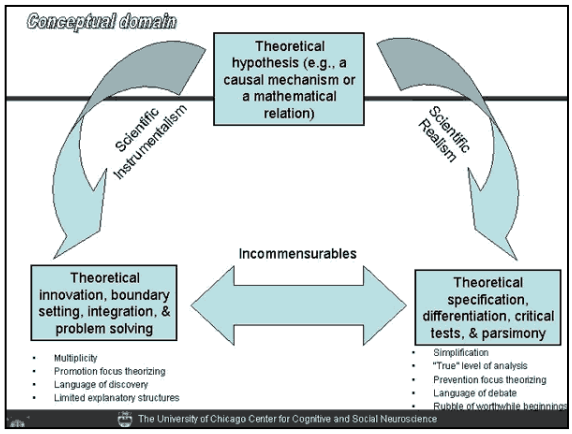
Scientific realism holds that scientific theories go beyond data to posit the existence of nonobservable entities – such as quarks, mental representations, and social cognition – which actually exist. According to scientific realists, the product of successful scientific research is knowledge that is independent of theory or

methodology (although theories may still be useful devices to organize this knowledge). Scientific theories grounded in the philosophy of realism attempt to describe the world as it really is –independent of human theories, perceptions, or measurements – and to establish what actually exists in it.

Scientific realism has a number of strengths that promote advances in scientific theory and method (Cacioppo, Semin, & Berntson, 2004). First, it contributes to the notion that a scientific theory should be falsifiable and subject to empirical verification. Second, realism presses for theoretical statements about phenomena one cannot observe by traditional empirical standards to be rendered observable by extending the range of our senses through the use of new instruments and procedures. That is, scientific realism underscores the importance of the development of new methods for theory. Third, scientific realism emphasizes the importance of a priori hypotheses, preferably embedded within a broader theoretical formulation, to guide empirical observation. Fourth, realism emphasizes that the purpose of empiricism in theory is to test whether the knowledge representation is true. When multiple explanations for a phenomenon exist, scientific realism maintains that at most only one of the conflicting hypotheses and theories could be true. Accordingly, scientific realism promotes theoretical specification, differentiation, warfare, and parsimony. Fifth, scientific realism promotes the notion that the explanatory power of a theory is defined by its predictive power. A theory among a set of theories that make the same empirical predictions may be preferred because it explains observable phenomena better than others – where by better one might mean generativeness and scope. Thus, a theory that produces a novel prediction is to be preferred over one developed later to account for the observation. Finally, realism encourages paradigmatic development that may be unique to adherents who are deeply committed to the Truth of their guiding formulation. Behaviorism in American psychology was certainly characterized by deeply committed adherents who pushed the boundaries of their guiding theories beyond what less committed adherents would have achieved.

Scientific realism also has its downside, of course. Scientific realism can be characterized as outcome driven in that it seeks to arrive at a conception that the theorist regards as the Truth; once discovered this Truth/outcome is defended against all competitors. The disadvantage of this is that it can impede progress and encourage defensiveness of and commitment to a theory well beyond its utility. That is, theoretical battles not only may eliminate the chaff in a scientific field, it may also savage worthwhile formulations that are less developed, more innovative, less vigorously promoted, or maintained by scientists with less political clout. As a result, realism can leave in its wake the ruins of worthwhile beginnings and the lost utility of theories that have fallen out of fashion. Newtonian mechanics remains a vital theory, because it remains a useful approximation In addition, realism promotes the idea that one level of analysis might be the right or true representation, thereby encouraging substitutionism rather than reductionism.

The perspective of scientific instrumentalism, in contrast, holds that the aim of scientific theories is not to discover Truth but rather to produce intellectual structures that provide adequate predictions of what is observed, and useful frameworks for answering questions and solving problems in a given domain. From this philosophical perspective, scientific theory represents convenient intellectual structures for predicting or describing in more abstract terms observable data, not actual structures in the world. In the view of instrumentalists, rational choice theory in economics, arousal theory in psychology, and Newtonian mechanics in physics each represent an important theoretical structure because it continues to provide a simple explanation for dealing with gross phenomena in its domain even though none is sufficient alone to explain what we now know about behavior in these domains.

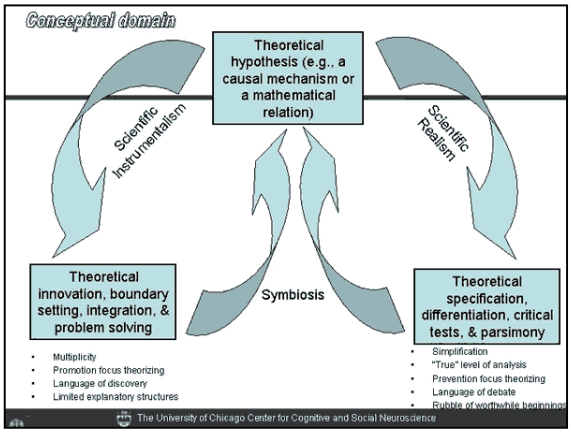


Whereas the perspective of scientific realism fosters theoretical specification, empiricism, verification, discrimination, and warfare between competing theories;

the perspective of scientific instrumentalism promotes open-mindedness, creativity, integration, consilience, and problem solving. This distinction is evident in the languages of science, which has been characterized as debate ("Victory!") or discovery ("Eureka!").

If the realist could be characterized as debater, warrior, and outcome driven (i.e., the search for and battles over universal Truths), the instrumentalist could be characterized as diplomat, discoverer, and process driven (i.e., the search for locally useful intellectual structures).

The strengths of the realist's approach is theoretical specification, empiricism, verification, discrimination, parsimony, and rigor, whereas its weaknesses are the zero-sum mentality about theory, include confirmatory biases, defensiveness, and oversimplification. The perspective of scientific instrumentalism, in contrast, promotes creativity, discovery, integration, scope, consilience, and problem solving, but at the expense of contest, precision, parsimony, and rigor. Both scientific realism and instrumentalism have advantages for the development and refinement of interdisciplinary theories spanning biological and social levels of organization. Scientific realism and instrumentalism represent such different approaches to scientific theory, however, that they have been viewed as incommensurable perspectives. The categorical distinction between scientific realism and instrumentalism may be overstated as a prescription for scientific practice, however (Cacioppo et al., 2004). Post hoc concepts, local microtheories, and pragmatic organizational schema will necessarily have an instrumental quality to them. We would suggest, however, that the ultimate goal should be to approach or approximate scientific realism. Some may suggest that neither is likely or even possible. But the instrumental quality of what could be called local science can be mitigated by expanding the generality, enriching the supporting empirical data base, modifying the theory to accommodate new or contradictory findings, and confirming bold predictions. Of particular importance in this regard, is the linking of concepts to other domains or disciplines—which have their own representations of putative reality. To the extent that such elaborated theories, representing multiple levels of organization and analysis, cohere with physiology, anatomy, sociology, etc, there is an increasing probability that it is pointing to some set of phenomena that have a lawful relation to reality, and that have generality beyond a local instrumental perspective. We may never really achieve a true scientific realism, but the pursuit confers great benefits on an instrumental approach, from the standpoint of generality, integration with other sciences, and the emphasis on the empirical rather than sociopolitical aspects of the scientific enterprise as the judge of theoretical and methodological merit. In this way, theorists might recognize the issues with and limitations of scientific realism without falling victim to a shallow instrumentalism.

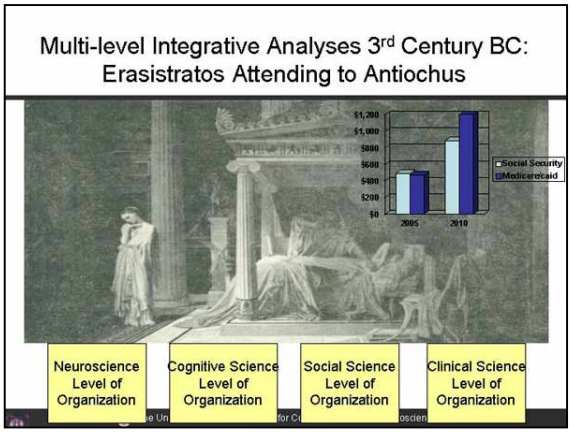


Our proposal, therefore, is quite simple: While philosophers may debate whether scientific realism or instrumentalism is best, as a practical lot scientists need not decide between the two but rather can capitalize on the strengths of each by adopting each perspective, in turn, in an iterative fashion, to guide theory and research. In the scientific method, the apparently incommensurable processes of induction and deduction are deployed in an iterative, integrative fashion to guide scientific theory and research. Within the conceptual domain of scientific theory, we propose that theorists should adopt the approach of a scientific realist and a scientific instrumentalist in an iterative, integrative fashion to guide theory and research. In the construction and testing of theory, there is a role for the warrior and for the diplomat, for the debater and for the discoverer, for these combined forces have the potential to enrich scientific theory and problem solving.

The tension between and iterative deployment of the perspectives of scientific realism and instrumentalism may be as integral to a useful, adaptable, integrative, and cumulative theory as is the tension between the

iteratively deployed processes of deduction and induction. In the proposed scientific symbiotic perspective, levels of analysis, paradigms, and theories no longer represent imperialistic empires but individual members of a community who are accorded station based on the specialized expertise, abilities, and problem solving capacities they bring to the community of science. Those members who are prosaic, redundant, or illegitimate fall to the rigors of parsimony, whereas those who are limited, quirky, but useful are accorded a place in the community.

In sum, the dichotomy between scientific realism and instrumentalism should be assigned to history, and the respective merits of each concept should be reevaluated with regard to the implications of each for how we think about, formulate, and evaluate scientific theory. The perspective of scientific symbiosis offers an alternative.



I'd like to end with the story of Antiochus and Erasistratos. Antiochus was the son of one of Alexander the Great's generals around the 3rd century BCE. Antiochus developed a mysterious malady that physicians were not able to identify. He showed pallor, weakness, and tachycardia under conditions that were difficult to identify under normal conditions of examination. Erasistratos, who already had a reputation for being a great diagnostician, was called to examine Antiochus. Erasistratos did so, but his approach differed from that of the other physicians who had attended to Antiochus. Erasistratos, as depicted in this slide, monitored Antiochus' physical responses during the course of a normal day, with special attention paid to how Antiochus' biological signs varied across social contexts. What he discovered was that Antiochus' weakness, cold sweats, racing heart beat, and pallor would arise whenever he was visited by his father's young, beautiful bride who was similar in age to Antiochus. Erasistratos diagnosed Antiochus' malady as love-sickness. The father of Antiochus resolved the problem by eliminating his new bride.

I share this story with you as a reminder that the influence of the social environment on biological processes and health has long been recognized, but this recognition has not been sufficient to incorporate social influences and processes into many medical or biological models of health and disease. The story of Antiochus is typically depicted in history not as the first documented case of love-sickness but as a moral tale.

With the maturation of scientific disciplines in the biological and social sciences, the development of new measures, methods and quantitative procedures that promote scientific analyses across multiple level of organization, and the pressing societal problems facing industrialized nations as their populations age and the economic burden of health care rises, an opportunity exists for new and exciting scientific advances in interdisciplinary fields such as biomarkers, which are bridging the abyss between social and biological domains. As illustrated in the story of Antiochus, multiple levels of analysis are synergistic in promoting an understanding of the health crisis facing industrialized societies. If we are going to make progress toward solving this crisis, me might take a lesson from Erasistratos.

Thank you very much.

(Applause)

McDADE: Are there questions, comments for Doctor Cacioppo?

HALTER: A couple things. One, you showed a curve kind of diminishing returns. In biology I think a lot of functions are sinusoidal, and you may be showing the top of the S; and the question is: How do you know where you are? You struggle at the bottom of the S real hard and finally maybe get somewhere where you can take off and then things plateau. You know, that seems to be --

CACIOPPO: I think you're absolutely right. In fact I know in my own research when I change and do something different, the first year, year and a half is difficult, right, because I'm learning a lot but most of my operationalizations and hypotheses are proving ill-conceived. I'm making all kinds of errors; and then things start to cohere and we make better progress. I stand corrected. The sigmoidal function would be a more accurate depiction.

HALTER: That relates to another issue. It has to do with I'm in a very big medical school, and we have stretched out our resources so thinly that everyone now is under intense pressure to support every minute. So where does the support come for that lower part of the S and the thinking time, retreats and all the intellectual challenge and stimulus that takes investment?

CACIOPPO: Yes, I agree. I think what you're seeing is one of the reasons many brilliant scientists stay in a particular paradigm too long. They may stay in the same paradigm their whole career because it is productive and easy, and they can avoid the high costs associated with changing paradigms. As a result, they have greater certainty of being able to pay for themselves 100 percent as they're being asked to do.

There are a number of ways to deal with this problem. One is to have more than one line of research going so that as you are enjoying the support from one line of research, you can incur the cost from beginning or rethinking others.

HALTER: Just one more kind of perverse point. The idea that one can fund oneself in one area that's kind of going well and use that to support yourself in another area that's taking time, that's also becoming much more difficult to being looked at with scrutiny by NIH, for example, and starting to look at people's time and doesn't want to support with one grant activities that may relate to another.

CACIOPPO: We have NIH people in the room. I think they should go back and talk about this at NIH because this is the most destructive, and if we eliminate that start-up support --

HALTER: We are now told to not charge time to writing our next grant on our current grant. We can't use that time. I mean this is a huge problem.

HEIMAN: You're not supposed to tell us.

CACIOPPO: If you sitting in this room agree that it is important to be able to have the opportunity to innovate in interdisciplinary scientific teams, then we have to start make the argument to NIH for doing so. In addition, I know of very few academicians who actually work only 40 hours a week. Changing questions or paradigms may take additional work for which one is not funded, and typically this requires more overtime work. Each investigator has to determine whether it's worth the additional time and effort. I personally believe it is.

WEIR: John, before I start, we're friends, right? I really love the Antiochus story in part because it taps into something that's always troubled me about disparity research from the perspective of (inaudible) and that is you've identify the source of the problem. Do you kill the young bride to make the problem go away? If you've identified the problem as, for instance, modern industrial society, do you have to make that go away to make people healthier? And so I guess what I'm inviting you to maybe offer is this kind of biomarker revolution if you like integrating biology with disparity research as giving us insight into pathways where interventions might occur short of radical social structure that (inaudible).

CACIOPPO: I think it's a very good point. We need to be cautious about moving too quickly to policy implications based on new scientific findings. Rather than sending the young bride away because of Antiochus' lovesickeness, might a bit more time and research have led to a cure for Antiochus and spared the bride?

SEEMAN: I'm curious from your experience if you could say a little more about how you think one can most effectively go about creating groups that allow you to do this integration of how you get together the right people and how you think about having an effectively group that does this integration.

CACIOPPO: First let me say I feel a little embarrassed to be standing up here talking about this when there are many other senior people here who have done this better than I. With that caveat, let me suggest one possible answer.

First, it is important to know in advance what each person is trying to achieve in his or her career, where they are in their career, and how they think about some of the issues about which we've spoken because these issues bear on whether they will be a good fit for how the interdisciplinary team is envisioned to operate. It is not just their expertise but their philosophy and where they are in their career that is important.

Second, I am a big believer in self-selection – to being explicit about how the interdisciplinary team is envisioned to operate and allowing people to participate or not depending on whether it is a good fit for them personally at this point of their career and research. I can't tell them, they need to identify the fit for themselves.

MELTZER: You talked about these learning plans for trainees. I was wondering if --

CACIOPPO: I got the idea of learning plans from you, David.

MELTZER: One of the things I was wondering if you could elaborate on some of the challenges faced by junior faculty sort of picking up on Jeff's question, particularly where you have different schools that have different expectations; and although they may aspire to sort of transcend these differences, it's often difficult, and sort of combined with that the issue particularly very busy, well intentioned but very busy, faculty members, sort of heavily mentoring people in departments that are not their own that may not always come to them initially with all the skills they need and so on.

CACIOPPO: It's a hard question, but my personal belief is that if a senior faculty member is so busy they can't mentor junior faculty, then they have missed one of the greatest joys their career offers. I had the good fortune of being told as an assistant professor that a full professor is someone who does more than just their research, they also serve as a constructive force in their classes, department, and science, and that one of the greatest

obligations was to promote the scientific development of students and junior colleagues. If one hopes to contribute bricks along which other scientists can step, one needs also to be concerned about the next generation of scientists.

Many sitting in this room have overcome numerous obstacles. Each has much to contribute so that their junior faculty need not go through the same obstacles. If everyone were to do so, it would be good for our institutions and for our disciplines. It helps advance the solutions to the problems we are addressing. Given the importance of finding solutions to the problem of rising health costs associated with chronic diseases in aging industrialized societies, we all have sufficient motivation to help junior faculty and students work effectively in interdisciplinary scientific teams. We also have to challenge tenure and promotion criteria that evaluate junior faculty only in terms of the number of single-authored publications they have achieved. Such a criterion is inconsistent with interdisciplinary research. We should instead consider evaluating junior faculty in terms of whether they have established a program of research and whether or not their program of research is making a significant contribution to science. This is, of course, what John Platt was saying in the quote in Slide 33 about bricks laying around the brickyard of science. His point is that lots of bricks isn't helpful. Having bricks that build something, that contribute something to the temple of science is what is important.

LINDAU: There was a slide that had symbiosis in the middle. If you could go back to it, my question pertains to that.

I think I can buy your argument that these two ways of thinking can be symbiotic, but I wonder two things; whether you think as a psychologist who thinks about brains whether some people tend to be more scientific realism brains and some people tend to be more instrumentalism brains. And a corollary to that is, do you think that the number of years one is a thinker relates to this, or do people tend to transition from one to the other? The one reason I ask this question is because Jeff's point that the emphasis in the early part of a career is on production not thought. I agree with you about the tragedy, potentially, and irony of that. If you're production focused, can you be a searching for truth person?

CACIOPPO: Well, I actually don't think what I'm suggesting is for people to be less productive.

LINDAU: I mean production in the number of papers.

CACIOPPO: The key is to be efficient. The likelihood that your hypothesis that you're pursuing is going to lead to something important is the joint probability that every step you took enroute to that hypothesis. If one has shown a confirmatory bias or published work too quickly, these probabilities will be lower than if there is no bias. The difference between these two strategies, in terms of the probability one's current hypothesis is likely to lead to an important insight, can become dramatic fairly quickly. A hypothesis based on three independent chance findings has only a 50% * 50% * 50%, or 12.5%, chance, whereas a hypothesis based on three independent findings for which there is clear evidence has, for instance, a 99% * 99% * 99%, or 97%, chance of yielding something worthwhile. As this illustrates, one can also be more productive, in terms of the number of studies published, with programmatic interdisciplinary research even if it takes a bit longer to satisfactorily complete any single study.

Are there individual differences in whether people think like realists or instrumentalists? I don't know, but I would think it likely.

As for differences across one's career, I believe you are correct in what you are suggesting: young scientists tend to be trained to think about science like a realist, whereas older scientists have sufficient experiences that they may be more likely to think about science in a nuanced fashion.

LINDAU: Can you see the person across the table as a realist versus an instrumentalist?

CACIOPPO: I believe so. We have all observed that among the most vicious people on NIH panels are young guns who want to demonstrate that they're smart so they rip apart an innovative proposal that, in fact, may have had real value. We have all also observed senior faculty on NIH panels who supported proposals that favored their theoretical perspective or level of analysis and did not support proposals that disfavored their perspective or level of analysis. It is not difficult to see these individuals are working from a position of scientific realism. Of course, others might take the attitude that everything is equally correct, that funding should be based on the importance of the question being asked, not on the quality of the theoretical analysis or methods that are proposed. Such panelists, I would contend, can also be deeply disruptive to a study panel's mission, and it is easy to identify such individuals (knowingly or not) as working from a position of scientific instrumentalism.

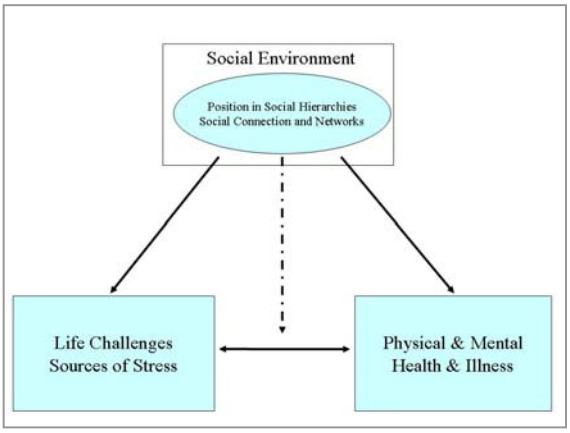
McDADE: Thank you so much, John.

Biomarker Data in Population-based Research: Conceptual Models, Analytic Strategies, and Hypotheses Tested

Maxine Weinstein, Ph.D.

Thank you for the invitation to be here. It's a pleasure to be in Chicago where it's cooler than our nation's Capitol which is enjoying August weather, hot, humid and hazy.

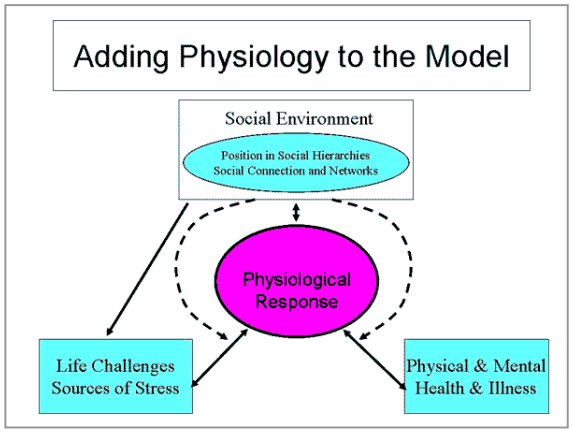
When I originally spoke with Stacy Lindau, she asked me to talk about some of our failures, and we have a few of those that we've made along the way. I'm fairly unembarrassed about talking about them because I think by and large we've had a lot of successes as well on this project.



I'm going to be talking about the Taiwan biomarker study; its acronym is SEBAS -- the Social Environment and Biomarkers of Aging Study. I'll briefly, briefly go through the project. I'll talk about some of the results, and I'll talk about how we are trying to make it a little better as we move into the second round of for us biomarker collection. Throughout I'll try to integrate comments about some of our successes and failures.

So this was the model that we started with. It's a very general model. The social environment affects exposure to stress. It affects physical and mental health, and all of these are reciprocal.

As you know, there are huge literatures connecting social environment and health, stress and health, but what we wanted to do was to add physiology to the model. We know that these factors of social environment, life challenge don't automatically result in changes in health; They have to operate through physiological pathways and mechanisms. What we wanted to do was unpack that black box, actually the purple circle, mauve circle in the middle, and that's what we set off to do. This is probably a very common kind of picture in the work everybody's doing here.



We did our work in Taiwan. In part this was because I had been working with people in Taiwan for years, since the early '80s when I was primarily interested in reproduction, but also because we were able to build on some wonderful work that had been done with what was at that time the Taiwan Institute of Family Planning and the University of Michigan initiated by Albert Hermalin. They began a longitudinal study of elderly in 1989 with major follow-ups about every three years. We had data in 1989, '93, '96 '99, and there's been another study 2003. In 2000 we took a random subsample of the 1999 participants and collected biomarkers. We will be going back this fall and starting in this fall to do a second round on those same people.

So we have a long series of the usual kinds of sociodemographic economic information that starts back in '89 and extends through 2003; and two collections of biomarkers in 2000 and 2006. I just want to put up a little information about Taiwan and the U.S. Although the culture is very different, in part it's interesting for those of us in the United States because of some of the similarities as well as the differences. You can see that overall expectation of life is very similar in Taiwan and the U.S. These data are for 2002. Cause of death is a little different 2002 for Taiwan, 2001 for the U.S. and major causes of death are not all that different.So at least in terms of -- if I can call it the ultimate outcome -- death, we're looking at similar factors. We're going to want to come back to this because surprisingly what we see is similar differences between expectation of life for men and women in both countries, about five and a half years. Even though we think that social position of women, at least traditionally in Taiwan was much lower than men, so it's a traditionally patriarchal society.

2000 Interview Information

- Update on health status (self-assessed, functional limitations, chronic disease)
- Update on CES-D, memory recall
- Update on histories (marriage, job, residence)
- **New topics**
  - Self-assessed stress/anxiety
  - Personal mastery (locus of control)
  - Recent crime experiences
  - Consequences of Sept. 1999 earthquake
  - Assessment of political situation
  - Relative SES position on ladder

Taiwan/US "Factoids"

	Taiwan	US
e_0 Males	73.0	74.5
e_0 Females	78.8	79.9
Δ (Male-Female)	-5.8	-5.4
Major Causes of Death (Rank)		
Cancer	(1)	(2)
Stroke	(2)	(3)
Heart Disease	(3)	(1)
Diabetes	(4)	(6)

Cause of death rankings reflect 2002 (Taiwan) and 2001 (US); expectation of life refers to 2002 (Taiwan and US)

Turning to our 2000 data collection, which is now done, and I'm happy to say the data are all available on the web. We sampled half of the PSUs that were in the longitudinal studies. We interviewed about 1,500 people and obtained biomarkers for about 1,000. We came very, very close to our objectives -- that was probably just good luck. We had a very high response rate for our interviews, about 92 percent among the survivors, about a 68 percent response rate for the physical exams. A major reason for nonparticipation was that they had just got a regular health exam and so they thought that there was no reason for a health exam. Some of them we screened out because we were afraid to bring them into hospitals. They were either too ill or frail.

2000 Data Collection

- **Objective**: sample half (27) of PSUs, interview 1500 (1/2 elderly), obtain biomarkers for 1000 respondents
- Actual numbers
  - 1497 interviews, 1023 with biomarkers
  - Response rate for interviews of 92% (among survivors)
  - Response rate for physical exam of 68% (among respondents): 61% elderly, 75% middle-aged
  - Major reasons for non-participation in physical exam
    - Just received or has regular health exams
    - R or family thinks R is healthy/no need for exam
    - Too ill/frail
    - Will be out of town/no time within scheduled days
    - Too much trouble

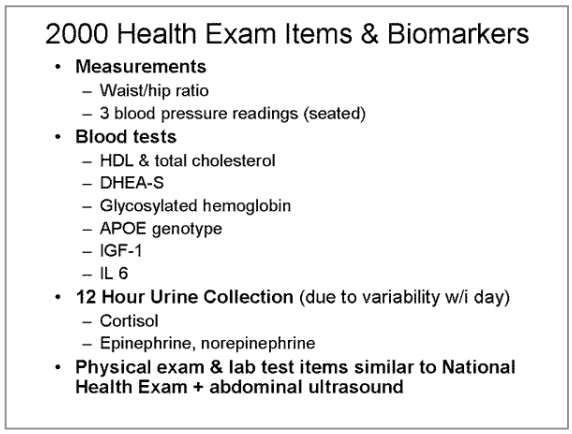
By the way, please interrupt me with questions.

LINDAU: I was just about to when you said that. So I think you answered it. The intention was for everybody, for all the biological and physiological data to happen in the clinical setting as opposed to in the home.

WEINSTEIN: Okay. So now you're pointing to one of our early mistakes. We interviewed the 1,500 (1,497 to be exact) people in the household, and what we lost there was a wonderful opportunity to nab as many biomarkers as we could. We got the biomarkers in hospital. The only biomarker we collected in household was the overnight urine specimen. So that was one of the things we learned, and this time around (to anticipate), one of our learning experiences is that we're going to try to collect much more in the household.

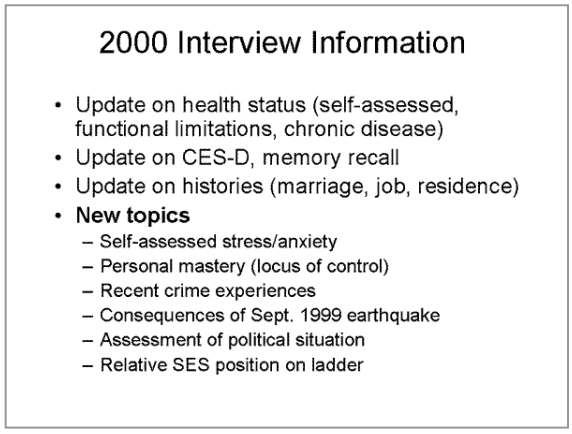
LINDAU: Thanks.

WEINSTEIN: Our new study will start will start with a pilot scheduled for this fall. What our overarching goal here is to reinterview and get new specimens from the people we got them from the last time around, and we're also going to collect new biomarker specimens from a refresher cohort. The study continues to bring younger people in so basically it's a random sample of population age 55 and above. We'll also be working to improve our household collection and to improve our methods.



Our household collection this time will include spirometry. We will be measuring blood pressure in home as well as in the hospital. We will be giving all our respondents collection kits for salivary cortisol so they will be collecting salivary cortisol in the evening before they go to bed, first thing in the morning, and then about a half hour after that. And we'll also try to do some functional assessments in the household.

That's actually very challenging. You get to see if somebody can walk a fixed distance and stand up or sit down from a chair, but in fact the apartments in Taiwan are very small. It's not easy to find the distance. And then, you know, what's a standard chair? So these are things that we have to work out as we move along. We'll be piloting in October, and I'm sure we're going to learn a lot from that as well.



The 2000 markers that we included were some simple anthropometry, weight-hip ratio, height, weight. We did three seated blood pressure readings. We got a fasting blood specimen in the morning in the hospital. We did the sort of standard blood test that's used by clinicians. We also (and this was mostly driven by work that had been done before, we were fortunate to have Teresa -- we are fortunate to have Teresa as a collaborator on this project, repeated a lot of the measures that were on the MacArthur study on the blood tests. We did a 12-hour urine collection from which we got cortisol, catecholamines, and then the rest of the hospital exam was very similar to the standard issued National Health Insurance exam for the Taiwanese that the elderly are entitled to have except we offered abdominal ultrasound, which turned out to be a very strong incentive for people to cooperate with our study.

In the interview information, we updated health status from 1999. We did CES-D, depressive symptoms, memory recall. We updated histories. We also added information that got at our overall objectives: additional information on stress and anxiety, personal mastery, locus of control, recent crime experiences.

You may remember that there was a major earthquake in Taiwan. I think, Teresa, was it 8.4? Very big. It came in the middle of our study and interrupted the schedule so that one of the reasons we had only a 68 response rate was that by the time we got to people, they had already had their National Health Insurance exam. On the other hand, it

provided a natural catastrophe, what can I say, a natural stressor, a particular kind of stressor but one that was very commonly felt. If you were on the island, that earthquake was big enough so that you felt it, and many people were affected. We also asked participants for their assessment of the political situation, and measured SES with a new instrument developed by Nancy Adler.

Here are some pictures. Nothing is simple. Taiwan weather is of course hot and, humid, so everything needs ice; packing the urine containers. I'll go through these quickly.

Processing urine specimens in hospital

Measuring waist & hip circumference in hospital

Blood pressure station in Fong-Shan Hospital, Kaohsiung

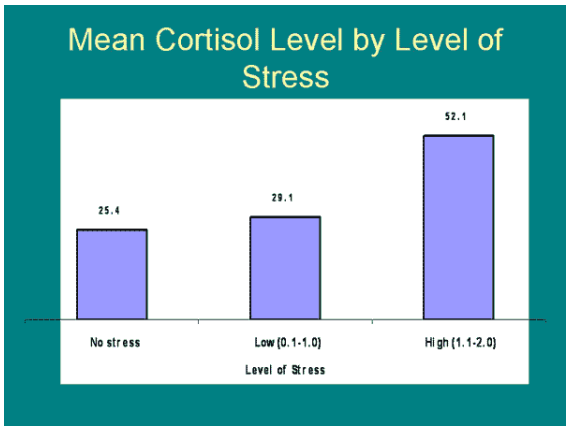
Hospital in Fong-Shan

Here's the blood pressure station in Kaohsiung; blood drawing station. I've whited out faces, but in fact people liked being photographed.

Measuring waist and hip circumference; processing urine specimens.

I come back to Thom's question about whether you really want a hands-on workshop, and I think the answer is yes, because I've whipped through these slides, but each one of these slides represents weeks of figuring out logistics. And here is the hospital in Fong-Shan.

Let's take a look at some of the results. This gets to what John was talking about, okay. If we assume that cortisol reflects stress, the good news is that it does seem that people



with no stress in their lives have lower levels of cortisol than people with moderate low levels and people with high levels.

We see the usual kinds of things that we would expect. Levels of functional limitations vary directly with levels of stress; socioeconomic status and health work in the ways we would expect inversely, and socioeconomic status and stress not as strong a relationship.

Yes, Jeff.

HALTER: Could you just define your stress measures?

WEINSTEIN: Yes. This is from a report to the Ministry of Health, and it's based on just three items about perceived stress, and I'll come back to that too because this was another learning experience.

One of the points that Thom made I think earlier was that we can use biomarkers to calibrate and validate self-reports, and so we looked at two measures here. One was hypertension; one was diabetes; and what we see is that self-reports of diabetes are pretty good, but hypertension is awful. If all we had were self-reports of hypertension to base our studies on, we would be missing quite a bit. These are -- one more comment. These are very comparable, and in fact this table comes from a paper in which we compared these values with the NHANES, and the numbers are not that different in NHANES as well. Yes, Jeff.

HALTER: My guess is your sensitivity for self-report for diabetes is overestimated because you have not used the same sensitivity of the test for diabetes as you're using for hypertension? You measure the blood pressure which defines hypertension. What did you measure for establishing –

WEINSTEIN: We had blood sugar, and we also had glycosylated hemoglobin, and in this case we used the glucose.

HALTER: Fasting blood?

WEINSTEIN: Yes.

HALTER: You're still missing 7 or 8 percent probably of people who have diabetes.

WEINSTEIN: Very possibly. I would never claim that these are perfect numbers up here. The big message here is the huge difference between hypertension and diabetes and probably one reason for that is that there was an island-wide service in Taiwan of testing for diabetes. People were much more well informed about their sugar levels than they were about blood pressure.

Yes, Colm.

O'MUIRCHEARTAIGH: I am a little suspicious of any condition that more than half the population has. What are the medical criteria for hypertension? How many people when we ask them think about what you mean?

WEINSTEIN: Good question. The self-reported question is: Have you ever been told by a doctor that you have hypertension? Now, recently the criteria have changed. I believe now that a stage II hypertension is defined as systolic above 160, diastolic above 100; but we did not ask that question. Bear in mind that this is an elderly population, so everybody here is above age 62.

HOUSE: What is your definition of the medical criteria, 140 over 90?

WEINSTEIN: For this we used 140.

HOUSE: Just one systolic over?

WEINSTEIN: Yes.

FREDMAN: How did you control for medication use?

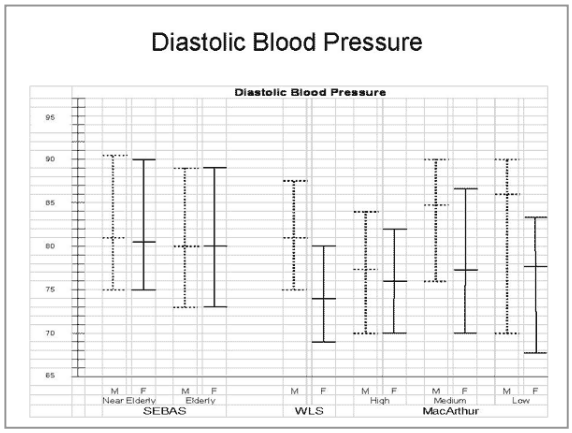
WEINSTEIN: We have that information.

FREDMAN: That could also affect the biological marker?

WEINSTEIN: It could, no question about that, and we've struggled with that in other analyses. To the extent that I can report analyses that I'm not really showing here, we've been surprised that, in fact, use of hypertensive medications has no effect on any of the analyses that we've looked at and so that's one of the things that we're looking into some more.

WEINSTEIN: So I just want to look at some simple results, and the main point to look at here are these male-to-female ratios; and you'll remember that I started off by saying that we expected that being female in Taiwan puts you at a disadvantage in terms of hierarchy; but for health, or at least for mortality, we usually expect a female advantage.

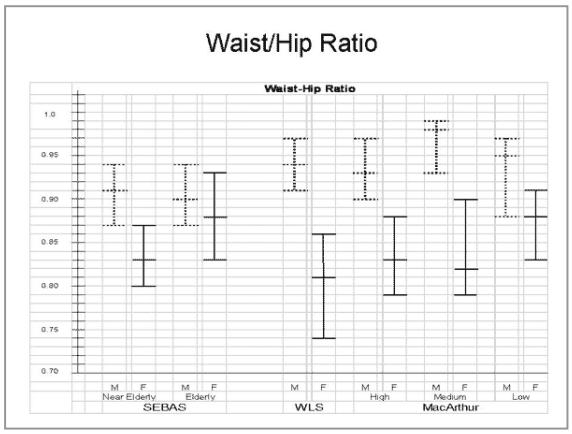
Indeed if you look at those male-female ratios in the United States over on the right-hand column, you'll see that death rates from heart disease and diabetes are uniformly higher for men than for women. In general heart disease is higher for men than for woman in Taiwan, but again in general the male-female ratio is much lower, and the pattern is reversed for diabetes. So in some instances in general what we see is that the female advantage is almost uniformly smaller, and in some cases it's reversed.



So we took a look at some of the biomarkers for this. What a picture, huh? What we've got here are the results of three studies using data kindly shared by the investigators. You'll see on the far right are data from the MacArthur study provided by Teresa and from

the Wisconsin Longitudinal Study provided by Bob Hauser, and for our Taiwan study SEBAS, and what we're looking at here is diastolic blood pressure.

The solid lines are female. The dashed lines are male, and you'll see that by and large on the right-hand side of the screen in the U.S. samples the top end of the female is much lower than the top end of the male. The medians are generally lower, but in Taiwan you don't see that, much closer distributions of diastolic blood pressure, and the same is true for waist/hip ratio. In fact when we look at almost all of the biomarkers we collected, what we saw was a much smaller female advantage relative to men.



So then we moved on, Jeff, to get to some of your questions, to looking more closely at stress other than these sort of big pictures; and we constructed an index of eight items based on the 2000 study of whether eight factors made you feel stressed or anxious and I want to come back to -- or maybe I'll just say it now. We didn't measure exposure to stress well enough, and that's an area where we're going to have to do a lot of improvement in 2006.

We measured it using some standard scales, not enough but some standard scales. We examined stress as of 2000. We looked at cumulative perceived stress over '96, '99 and 2000; and we looked to see how these stress indexes were related to allostatic load. We did allostatic load a little differently from the original formulation, and people were going back and forth about the best way to formulate it. The good news is that at least in this paper it was pretty robust to different formulations, but I'm perfectly open to quibbles about that.

Perceived Stress

- Perceived Stress Index (2000)
  - Whether 8 factors "make you feel stressed or anxious"
  - R's life: own health, financial situation, job, getting along with family members
  - R's family: family or child's health, financial situation, job, marital situation
- Cumulative Perceived Stress (1996, 1999, 2000)
  - Constructed to be as similar as possible using 3-5 items

So we looked at the effects of stress on this allostatic load index. What you'll see is that females tend to be more stressed than males. When we look at either current stress or cumulative stress and its relationship to allostatic load, we see a strong relationship. There is an interaction effect with current stress and being female, but all of them are strongly related to our measure of allostatic load, perceived stress.

However, when we tried to go back and look at cumulative stress by using the early longitudinal studies in measuring stress across waves and so forth, we find only weak relationships between stress and allostatic load; and as I've been thinking about what we're finding, I think that we need to rethink the way we frame this in a couple ways.

First is that -- I mean this sounds so obvious, there are all different kinds of stress. There's psychosocial stress. There's physical stress. We didn't do a good enough job measuring each of those domains.

Second, each of those domains may play out physiologically in different ways. Some kinds of stress will be reflected in, say, SNS activity. Others will affect HPA activity; and our measure of allostatic load mixes them all together. That's a good thing in some ways because we might miss particular stressors -- particular biomarkers that you expected to see cumulatively. On the other hand, we might be in a situation where we're masking out these different pathways.

Cumulative Physiological Dysregulation

	Mean or percent	Standard Deviation	10th percentile[2]	90th percent
Biological markers				
Epinephrine (µg/g creatinine)	2.6	2.6	B.A.S.	5.6
Norepinephrine (µg/g creatinine)	21.9	9.9	11.2	34.7
Dopamine (µg/g creatinine)	221.9	1110.2	87.4	226.7
Cortisol (µg/g creatinine)	28.6	52.6	8.7	48.0
IGF-1 (ng/ml)	105.1	48.1	53.1	168.0
IL-6 (pg/ml)	1.8	8.2	B.A.S.	3.4
DHEAS (µg/dl)	81.2	59.1	20.9	152.4
Systolic BP (mmHg)	138.4	20.6	114	166
Diastolic BP (mmHg)	82.2	11.1	70	97
Total cholesterol (mg/dl)	200.7	39.6	153	252
Total/HDL cholesterol	4.4	1.4	2.8	6.1
Triglycerides (mg/dl)	123.4	91.0	54	204
Fasting glucose (mg/dl)	107.1	37.9	84	138
Glycosylated hemoglobin (%)	5.8	1.4	4.8	7.1
BMI	24.4	5.6	20.0	28.9
Waist-hip ratio	0.88	0.07	0.80	0.96
Cumulative physiological dysregulation score[3]	3.4	1.8	1	6

"Perceived Stress and Physiological Dysregulation in Elderly Humans," Goldman et al. 2005 Stress 8(2):95-105.

Final question that we've just recently started looking at is mortality follow-up three years later and the biomarkers in 2000, and you can see that for the dark black line are what we call preclinical markers, for want of a better word, --the epi, norepi, cortisol, DHEAS. It's the stuff for which there are really no standard clinical cutoffs at this point. So what we have here is that yes, indeed, the preclinical markers do a better job at any level of specificity. They're more sensitive and vice versa than the standard markers.

Every study is the pretest for the next study. We need better questions about exposure to stress, and about trauma-- something that we hadn't even thought about the first time we went into the field. We are working on getting better measures of personality states and traits: How do you cope with stress? Are you an anxious person? Are you an optimistic person?

Coming back to Stacy's question, we'll collect more information in the household. We'll be doing functional status, blood pressure, salivary cortisol; and we will be looking at some different biomarkers. We'll be doing heart rate variability in the field, measuring C-reactive protein, a couple other things, measuring telomere length.

EVERY STUDY IS THE PRE-TEST FOR THE NEXT…

- Better measures of exposure to stress and what about trauma?
- Better measures of personality states and traits
- More information collected in the household
  - Anthropometry
  - Functional status
  - Blood pressure
  - Salivary cortisol
- Some different biomarkers
  - Heart rate variability
  - C-reactive protein
  - Telomere length

I want to mention my wonderful collaborators especially from Taiwan, Chang Ming-Cheng, most people here will know him as Jack Chang, Yi-Li Chuang, Yu-Hsuan Lin and I-Wen Liu, who made this study really possible. I think the other thing that made this study so wonderful, and again this fits in with John's comments earlier, was that we all really like each other so that when we had to hang around each other, there was no hostility. We were able to dig in and argue very constructively.

One advantage of not being in a department at Georgetown is that I get to work with anybody, and I'd like to mention that we have some relatively new collaborators as we've expanded our work. So Paul Aisen, a neurologist, has joined us working on cognitive function and Alzheimer's. Diane Yeager, who is in the Theology Department has started working with us on linkages between religion and health. Jim Baraniuk has started working with us on piloting some studies of peptides, and a former student , Tristan Gorrindo, now up at Mass General has been working with us also on some blood pressure and cognitive work, so it's been a wonderful experience in that realm.

Hsieh Hsieh.

MAHAY: Anybody have any quick questions while we switch over?

FENDRICH: (Inaudible) Any medications for people or herbs or anything that may affect some of the measurements you're taking also just to generalize it?

WEINSTEIN: We do have self-reports of medications. Those are notorious. I mean if you taught with anybody who really knows what they're doing, they'll tell you have to go in their home and look in their medicine cabinet and see what they actually have; and we don't do that.

We also do ask about use of traditional medicines, but again those are self-reports; and this could get to part of the answer to your question, which is we don't see much effect, and maybe it's because we have so much measurement error.

MAHAY: We have time for one more question.

CRIMMINS: That was very interesting. Can you tell me why you chose to do abdominal ultrasounds, what you were looking for and what you got out of them and are you doing them again?

WEINSTEIN: The abdominal ultrasounds were proposed by our Taiwanese colleagues as a mechanism for getting cooperation from the participants, and I can only tell you that it worked perfectly. We had almost 70 percent cooperation, which is a very high participation rate for dragging people into hospitals.

CRIMMINS: Now, what did you get out of them besides we showed up?

WEINSTEIN: It was strictly a service part of the protocol.

LINDAU: You didn't collect those data or analyze the data?

WEINSTEIN: We haven't analyzed those data in the sense of looking at them in our models. They were provided back to the participants.

CRIMMINS: With diagnostic advice or something? You didn't give them their picture?

WEINSTEIN: No, no, we didn't. We did have them read by physicians.

CRIMMINS: Looking for tumors?

WEINSTEIN: I don't know Hepatitis is common in Taiwan; the ultrasounds were used for liver problems.

CRIMMINS: It seemed an expensive incentive.

LINDAU: Incidence of incidentalomas, that's what we call them in the medical world, would be you would think quite high. It lead to surgical interventions or other anxieties about -- it's very interesting.

WEINSTEIN: People wanted this. That was why we did it.

MAHAY: I think it will be an interesting discussion for dinner. Thank you very much.

Next I'd like to introduce Doctor Sastry. He is a senior social scientist at RAND and the associate director of the labor and population program there. He is the principal investigator for many ongoing projects including neighborhood and family effects on stress and health and models of birth weight and low birth weight. Thanks very much.

Integration of Social and Biomarker Data in the Los Angeles Family and Neighborhood Survey

Narayan Sastry

I will be talking today about the Los Angeles Family and Neighborhood Survey (L.A.FANS), which I co-direct with Anne Pebley from UCLA. There is a large number of people involved in this study, and has been for a number of years. I am not going to acknowledge each of them, but their help and assistance has been greatly appreciated over the years.

Today I am going to give you some background about the L.A.FANS study: the goals of the study, the design, and the content. We have completed one wave of the survey and are currently planning the second wave, which will go into the field by next year. I will tell you about our plans for collecting biomarkers in Wave 2 of L.A.FANS: the specific measures, the plans, and so forth. I am not going to be talking much about analytical strategies, conceptual models, or hypotheses to be tested because we do not have any results yet.

The main goal of L.A.FANS is to study the multilevel effects of neighborhoods, schools, families, and peers on the health, development, and well-being of children and adults. A closely related goal is to understand patterns of residential mobility, segregation, and neighborhood change: where do people live, why do they live where they live, how do they move over time, and what factors shape these decisions? The reason for studying those two processes in a single study is to advance research on neighborhood effects.

There are three specific ways in which L.A.FANS is advancing neighborhood effects research. One is to collect richer data about neighborhoods — that is, more information about social and other processes in these neighborhoods. Another is to collect better data on how people perceive their neighborhoods, how neighborhoods are defined, how they are operationalized, and how neighborhoods compare to the social space where people live -- where their children go to school, where they shop, where they worship, and where they work. Being able to look at these factors together allows for better measures of how the physical neighborhood compares to where people spend their time. Finally, the ultimate goal is to investigate the issue of endogenous neighborhood effects and to assess the degree to which neighborhoods have a causal effect on health and other outcomes.

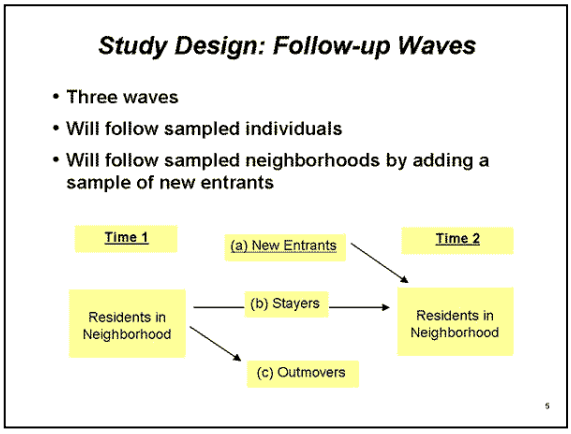
We know from many observational studies that neighborhoods can affect health. But it is also the case that health can affect neighborhood choice — that people may choose where they live based on their health status. Alternatively, other factors — measured or unmeasured – may shape people's choices about where to live and their health outcomes. Collecting better and richer data about factors that affect people's choices about where they live and their health outcomes — and to collect this information dynamically so that one knows how these processes unfold — is key to advancing research on how neighborhoods not only affect health but a variety of other outcomes as well. Lastly, since this work is based in Los Angeles, there is considerable diversity along dimensions of income, ethnicity, and race that is captured in the survey.

A quick overview of the sampling design of L.A.FANS might be useful. For sampling purposes, we operationlized neighborhoods as census tracts. We selected 65 census tracts in Los Angeles County for the study, from a total of about 1,650. The tracts were stratified by poverty status, and poor and very poor neighborhoods were oversampled. Within each sampled neighborhood, we selected 40 to 50 households for

the study. Our sample of 3,000 households is highly concentrated geographically, which allows us to develop measures for these specific neighborhoods. In particular, it was easier to collect data from administrative sources or as part of the fieldwork to characterize these neighborhoods.

L.A.FANS sampled two individuals – an adult and a child – from each household and and we follow those two individuals over time. We also interview other respondents, in addition to the sampled individuals. The household head provides information about the household's economic status. Because many of the children are quite young, we interview their primary caregiver to learn more about their living circumstances. Finally, we interview a sibling of the sampled child to be able to disentangle family effects from neighborhood effects.

One key innovation of the study is that it is a panel not only of individuals, but also of neighborhoods. We follow the sampled individuals over time, whether they stay in the sampled neighborhoods or move out to look at the factors that shape why people leave their neighborhoods. We also sample new entrants who moved into the neighborhoods, to look at the reasons why people move into these neighborhoods. There are also people who have stayed in the neighborhood. Together, these samples gives us a picture at both time 1 and time 2 of who actually lives in the neighborhood and allow us to characterize the process of neighborhood change.



The first wave of L.A.FANS was conducted in 2000 and 2001. We interviewed about 3,000 households, 3,600 adults, and 3,200 children. Response rates compare favorably to other nongovernment social surveys.

The L.A.FANS sample reflects the diversity of Los Angeles. It includes a large number of Hispanics, and many of the interviews were completed in Spanish. There are, however, a relatively small numbers of blacks and Asians in the sample. Los Angeles is one of the prime destinations for immigrants to the United States. That fact is certainly reflected in our data, with over half the sample comprised of first-generation immigrants.

The Wave 1 L.A.FANS data are publicly available, and can be downloaded from the L.A.FANS website which is hosted at RAND. The more interesting analyses of the data require the neighborhood characteristics, for which a restricted data use agreement is needed. The restricted data provide information on sampled tracts in the study, which is otherwise not released due to concerns about deductive disclosure of respondents' identities.

In the first wave of L.A.FANS, only self-reported health measures were collected. Information was obtained about self-assessed overall health status and reports of chronic disease and of height and weight. These data have not been used extensively in part because of problems with self-reported data.

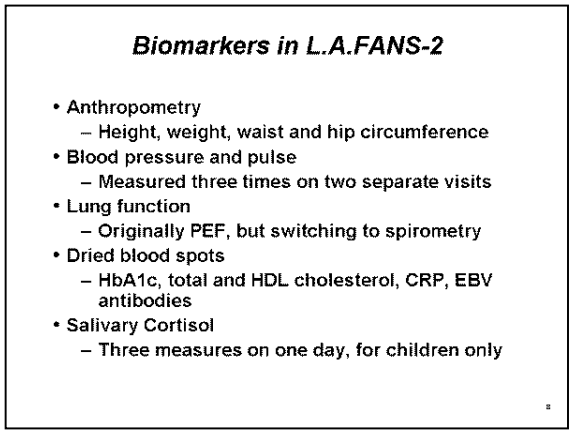
In analyses of self-assessed health status, one of the interesting findings to emerge was that the language of interview had a large effect on reported health status. L.A.FANS has a large sample of Hispanics, and about half of them were interviewed in Spanish and the other half in English. The language of interview apparently affects how people respond to questions about their health status. The problem is that the translated health categories are never exactly comparable. And there may also be cultural factors that shape people's choice of the language for the interview and how they think about their health. This is a tough issue to address. However, in Wave 2 of L.A.FANS we are administering a set of vignettes to calibrate self-reported health status to get at some of these issues of differential reporting and how it is related to individuals' characteristics.

The reports of chronic disease, as you saw from Max Weinstein's presentation, really do not provide much information without biomarker or other information to know who actually has the disease. This is because undiagnosed disease is an important issue, especially in a population such as this. Finally, height and weight, which is perhaps our best measure of health status, is also difficult to get from self-reports.

The data from Wave 1 that have been used most effectively are the objective measures, such as assessments of children's academic skills. A lot of time and effort was devoted to conducting reading and mathematics assessments of children and their primary caregiver. The standardized results provide a great measure of children's achievement in school. In Wave 2, we want to collect similar types of objective measures of health status.

Wave 2 of L.A.FANS is funded by four different grants from NIA, NICHD, and NIEHS. Work on the project began in mid-2004. We have a pretest planned for late 2005. Our goal is to be in the field with the main survey by early 2006, and we have a 20-month field work period planned. Thus, it is going to be a while before any data are available from the study. But when the data have been collected, they will all be released as public use and restricted use data.. The Wave 1 data from L.A.FANS were released about three months after fieldwork ended and we expect the release of Wave 2 data will occur as quickly.

The fieldwork for Wave 2 of L.A.FANS will be conducted by RTI. The health measures will be collected by a group called EMSI, which performs health insurance assessments and has assessors around the country with a pretty dense network of them in Los Angeles.



In Wave 2, the health measures we plan to collect include: anthropometry (height, weight, waist and hip circumference); blood pressure and pulse (both on the initial visit by the field interviews and again when the health technicians visit); and lung function. Originally, we were going to measure lung function by peak expiratory flow, but have switched to a more comprehensive measure of spirometry which is no more difficult to do in the field but requires much more expensive instrumentation which we are in the process of obtaining.

We also plan to collect dried blood spots to assess the standard group of measures: hemoglobin A1c, total and HDL cholesterol, C-reactive protein, and Epstein-Barr virus antibodies. Finally, we plan to assess salivary cortisol for children. Currently, we are still in the process of developing the protocols for collecting the various objective health measures.

The goal of collecting biomarkers in L.A.FANS-2 was to obtain objective measures of health status and to be able to study actual and emerging chronic disease at all ages, including among children. Although some biomarkers will not be collected for the very youngest children, other measures, such as height and weight, will be collected for everyone. A related goal is to collect wherever possible both self-reports and objective assessments to examine the issue of untreated and undiagnosed disease.

The key chronic diseases we focus on for adults are conditions such as diabetes and hypertension. For children in Los Angeles, asthma is the key issue. Conducting a full spirometry assessment is far better than collecting peak expiratory flow for assessing asthma symptoms among children. Another objective is to look at the precursors of chronic diseases — such as cholesterol levels, high blood pressure, and physiological markers of stress — in order to examine how social factors and neighborhood characteristics affect stress and early onset of chronic disease.

Documenting and investigating racial, ethnic, and socioeconomic status disparities in these measures is another key area we are investigating. We have interesting diversity in our sample. There is a large number of Hispanics in L.A.FANS—large enough to allow us to look at a number of important distinctions and differences such as differences in length of time in the U.S. and in the place of origin (distinguishing, for example, immigrants from Mexico and from Central America).

To summarize, the main research questions that the health measures in L.A.FANS are designed to address include: patterns of undiagnosed and untreated disease, family and neighborhood effects on health (a key strength of the L.A.FANS study), health status of immigrants, and disease management and treatment.

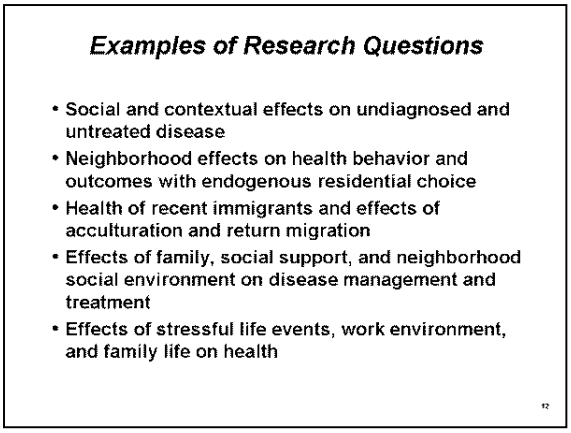
One of the strengths of L.A.FANS is the detailed information about where respondents live. We have geocoded household locations, as well as respondents' places of work. Because we are working in a single location, we are able to collect detailed information about sources of health care and available health facilities. Thus we are able to look at physical access to care that goes beyond just as-the-crow-flies distances to look at network-weighted distances. For example, we can look at public bus routes and how road networks and public transportation affect access to care, and how that, in turn, affects treatment for disease and disease management. We can also measure effects of life events, work environment, and stress on health.

A practical consideration that has emerged as we begin to implement our plans for Wave 2 of L.A.FANS is maximizing response rates. The population we are studying includes a large group of recent immigrants, many of whom move on to other places or return to their country of origin. Response rates are a challenge for most surveys, in the sense that surveys want to maximize response rates and minimize loss to follow-up. Response rates are not directly related to the biomarkers, except that over time higher non-response rates will result in a potentially more select sample. Minimizing overall nonresponse — and, in particular, nonresponse to the biomarkers — is a major goal for the study.

We will collect the blood-based biomarkers in a separate visit from the household interview. There are several reasons for this. Foremost is that in California we are told that you need a health professional (e.g., a nurse or phlebotomist) to collect blood samples, including those obtained through a finger stick. One alternative is to ask the respondents to perform the fingerstick on themselves. However, we cannot ask a field interviewer to perform a fingerstick unless they are trained and certified. The requirement to have a health professional means that a separate visit is needed. We can only schedule this visit after the household interview is complete, and obviously the health professional is not necessarily going to show up right away. We need at least 48 hours to schedule their visits. Once you have introduced that time lag, there is a potential for reduced response rates. Consequently, we cover as much as we can during the main interview - everything except the blood spots are done by the field interviewer. This was not our original plan, but it makes more sense.

A related issue is reducing respondent burden. We ask the interviewers and the families to do a huge amount already. This is a multipurpose survey – and everybody's got their questions in there. In Wave 1 of L.A.FANS we had two types of households: one with just adults and the other with adults and children. On average, the interview times were about an hour and a half in households with just adults and about four and a half hours in households with adults and children. This includes interviewing up to two adults, interviewing up to two children, as well as doing reading and math assessments. Adding the collection of biomarkers at the end of the interview raises the issue of respondent fatigue, which is a challenge. One factor working in our favor is that these visits are actually spread out. The interviewers do not sit there for four and a half hours on any particular day. Rather, completing the interviews often involves multiple visits. So the burden is reduced a bit in that regard. Also, the health results are potentially of interest to respondents, especially with a population that is poor and lacks regular access to health care. We hope that respondents derive a real benefit from having the results of these health tests.

We are currently dealing with a number of practical questions and issues and are looking for advice and guidance. People here have dealt with a lot of these issues in their current research and we are interested in hearing about how they dealt with them, getting suggestions on protocols and procedures, storing samples for adding additional tests later, and information on what new tests are emerging. One thing that we are particularly interested in is protocols for assessing validity and reliability of the test, retest procedures, and validating lab procedures. I will leave these as open questions that I would appreciate answers for at some point, given where we are in this process.



JORDAN: I have a question. I'm curious as to the choice of EBV. What was your thought behind looking for that?

SASTRY: Thom McDade suggested it, and we thought it was a great idea. The idea behind the EBV antibodies is that they are a marker of physiological response to stress, that is triggered under stressful situations. Thom can give you --

JORDAN: You're looking at reactivation?

SASTRY: Yes.

McDADE: It's a measure of your ability to control a latent infection. It's the best, most well-validated, immunologic measure for elements of chronic stress, far from perfect but the best we can do with our blood spots.

FENDRICH: What's the age range of the sample?

SASTRY: Nobody is excluded based on age. Kids zero to 17, and we sample adults. We have everybody.

FENDRICH: You're not doing any toxicology stuff. Are you looking at drug abuse in all this training that you're doing? Are you screening for drug abuse in any of your biomarker testing?

SASTRY: No.

FENDRICH: Why not?

SASTRY: With the dried blood spots or with sort of more generally?

FENDRICH: Urine? Looks like you're doing something with oral fluid.

SASTRY: We don't have urine.

FENDRICH: But you have oral fluid.

SASTRY: For kids, yes. We considered looking at smoking, but I think at this point it's unlikely due to budgetary restrictions. We have some of these samples. If we wanted to add more tests, we need to go out and get more money to do them.

HOUSE: There are potential issues in terms of human subjects considerations if you're asking about things that might reflect a (inaudible) behavior, and there's also the potential for a really chilling effect on response rates if you are measuring anything that could reflect adversely or be used adversely against the person, so I think there's been a tendency to stay away from that in large population studies for those reasons.

SASTRY: We ask people about it. We ask kids and they have a self-administered --

FENDRICH: You're asking people survey questions, but I mean with oral fluid my experience is if you ask people to participate in oral fluid screening for drug use, you get about 90 percent who will do it in a general population. I can see it maybe changes the nature of the study but if you really want to understand some of the risk factors.

SASTRY: Yes. There are other things that people have asked about and wondered why we weren't addressing -- issues like sexually transmitted diseases. Those are valid and interesting things to study. I think we circumscribed our study to focus on the particular issues of stress and emerging chronic disease.

ADAM: As a follow up to that, cotinine might be important as a control variable for cortisol. There's science now that secondhand smoke can influence cotinine levels in kids which in turn is related to their cortisol levels. We should talk also about what you're doing with cortisol with kids of different ages because we need different assessment protocols, and the (inaudible) rhythms change with age as well, so talk about that.

GARFIELD: This brings up an interesting point of where community falls into it. I know at Northwestern in Evanston we're working with UCLA on a community child health network which has a strong community component to it, and I'm wondering how you guys have engaged the community with that since a lot of the studies that you're doing, a lot of samples that you're drawing, could be controversial; and I think Maxine said the thought of nabbing as many biomarkers as possible is exactly how I think about it too. The conditions you get in the blood and let's get as many things as we can. The community doesn't see it that way.

SASTRY: Yes, yes. That's a good question. We've had a lot of community interest in this study. But we face one real problem, which is working with the specific 65 communities. The nature of the data and what's available means that it's really tough to release anything on a particular community. To have that kind of involvement where these results feed back and are publicized for a specific neighborhood is a problem in neighborhood of this size because you can really figure out almost immediately who the respondents are based on their characteristics.

So although we've had a large amount of interest and specific neighborhoods wanted to know if we have sample tract in their local planning area, we've actually refused to provide that information. The way that we've worked with communities is through larger L.A.-wide entities such as First Five L.A. (the Children and Families First Commission, funded by Proposition 10), by working with the county health department, and also by disseminating our results. There's been a fair amount of effort to publicize the findings—for example regarding school readiness. But in terms of actual involvement in communities, it's really tough.

WALLACE: Are you making any environmental measurements?

SASTRY: Yes, I think so. One of the things that's emerging now is that there would be these measures of asthma and doing a full spirometry. We actually have a group of collaborators who are interested in exposure to air pollution and how it's related to kids' asthma.

They have a funded proposal that goes hand-in-hand with this to install air pollution monitoring equipment at a sample of households in each of the 65 neighborhoods. It builds upon the network of air pollution monitoring stations in L.A. which is useful but is not dense enough to characterize the fine differences between the neighborhoods in air pollution.

In terms of environment, we also collect a lot of information from interviewer observation. Systematic social observations are being done here in Chicago where they characterize more the neighborhood social and physical environment, not so much environment in terms of air pollution, but instead in terms of housing quality, graffiti and trash on the streets, things like that.

WALLACE: There may be some simple things like indoor temperature, for example, people who don't have air conditioning.

SASTRY: Yes, yes. We collect some of that as a component of the home inventory. We don't ask about temperature but about the environment and how cluttered it is, whether it's clean, whether it's dark, whether it's well lit, so temperature could be a good one.

ADAM: Just to throw another one into the list, noise.

SASTRY: Yes. This is something I think you would want to assess over time, which is what the air pollution study would do. They would actually drop off these monitors and collect them after a couple years.

LINDAU: Thank you for an excellent presentation. I've been curious about the study, and I know a lot more about it now and I think it sounds very exciting. With regard to the biomarkers, which will you report back to respondents, which will you not, and what work are you doing if at all to anticipate the cultural meeting or response? You have such a diverse sample for the different biomarkers you're collecting.

SASTRY: We're reporting back the ones that have clinical significance, not the EBV antibodies but certainly even some CRP, so whatever the standard protocols is for reporting back, that's what we plan to do.

LINDAU: Do you have people call in for their results, or how will you notify them of their results?

SASTRY: By letter. We'll send them a letter. That's our plan. So we have some of the stuff that we can tell them right away, height and weight, and other. Blood pressure is somewhere in between the stuff that requires testing.

LINDAU: Have you thought about giving respondents an option as to whether or not they want their results sent to them by letter?

SASTRY: As opposed to not sending them at all?

LINDAU: As opposed to not getting them at all.

SASTRY: No, actually we haven't thought about that.

LINDAU: Or not by letter, some because of low literacy rates or a language barrier.

SASTRY: We do everything in English and Spanish. All our respondents, in order to be selected for the sample, have to speak one of these two languages. This is the reason why we have Asians comprise only 7 percent of the sample and it's a little bit of a curious Asian group for that reason. Literacy, it could be an issue. It could be an issue.

LINDAU: I think it very likely will be an issue, and I think I would recommend considering giving people an opportunity to opt out of getting their results. They may not want that.

SASTRY: Or getting them in another format.

LINDAU: As another alternative.

SASTRY: That's a good point.

HALTER: I have a biological analytical question, and maybe Maxine could comment too. Biological markers that you have have kind of different time constants associated with them. Your blood pressure is changing every minute, seconds. Your body weight is a pretty long-term measure. Your glycosylated hemoglobin is an average over weeks to months. Cortisol of course fluctuates during the course of the day. So I'm trying to analyze and relate biomarkers to life, social stress, living situation. How do you build in the time constant measure into the analysis?

SASTRY: That's an excellent question and something we've thought a bit about. I guess it's something that's ultimately best addressed on the research side. We're collecting all these data right now and in that context the best way that we have to deal with the issues is by the types of data we collect. The ideal circumstance would be to collect more data over more regular periods of time for the measures that change rapidly, and to some extent we do that with blood pressure. We have three measures each at two separate points in time at least two days apart, maybe longer. We'll get some sense of changes over time but obviously not perfect.

We do collect a detailed and extensive event history data that goes back in time between the two waves of the survey, and even further back in time before the first wave. It doesn't address the issue of timing, but it provides a greater set of options for people to deal with that. You're not going to get around the problem that you mention. When these data are collected, they'll be cross-sectional, but you can characterize people's circumstances like where they live, for example. So we have a residential history, a continuous residential history that covers the last seven years for people in the sample. So I don't have a good response to that.

HALTER: Maxine, how do you deal with that?

WEINSTEIN: You put the question that I raised in the end in a slightly different context, and some of these studies are using this allostatic load framework which is itself an assumption; that is that exposure to stress over time will result in these biomarkers being driven sort of out of normal operating ranges.

So the underlying assumption is that even though you're getting blood pressure today, it reflects cumulative studies. That's one of the reasons that I find our nonresults – the nonconfirmatory results - so very troubling. When we go back and we look at cumulative studies over a long time, we're not finding strong relationships.

So I don't have a good answer to that. We have two hopes on that. One is that the second round where we're able to measure change in a five-year period and relate it to changing experience over that same five-year period will help shed some light on what's jointly driving those.

The other is with this more general peptide analysis where we'll be looking at a higher range of biomarkers that kind of reflect potentially longer term measures; that is how you are metabolizing the proteins that may be more reflective of normal range processes. I don't have the answer.

HALTER: There are activity monitors, blood pressure monitors that one could get blood pressure averaged over the week, and so I mean there is some technology involved but it's not --

WEINSTEIN: We did halter monitoring in a study I've been involved in Moscow and St. Petersburg, and again the results are not entirely clear. Now that's only 22, 24 hours battery life. So it's really 22 hours. But you do see changes. So your question is I think even more pointed. I don't have the answer.

SASTRY: I suppose the best response at this point is to try to collect all the information as best you can, the types of things you think might affect that outcome. If you're collecting blood pressure at a particular point in time to find out the person's circumstances; did they just eat? Did they exercise earlier that day?

HALTER: I'm also wondering if there's an analytical approach to this in terms of different measures at a different time, constants that are built into them. Genetic analysis is what you have your whole life.

SASTRY: Exactly.

WEINSTEIN: That's not that simple either here.

CRIMMINS: No, that's not simple anymore.

HALTER: Well, you may express it.

SEEMAN: I just wanted to add something in response to Jeff's question. I think this is the chronic sort of tussle we have between trying to get more detailed information that you could do on smaller numbers of people and then these large population studies where we have issues around just our own logistics and feasibility as well as how much you can afford to do on people because of the cost. We realize for any given individual we're not getting a perfect picture of what goes on; and largely most of the protocols we try to employ are designed to get at what we think hopefully is a snapshot of a basal level so a blood pressure reading isn't meant to reflect what goes on across the day but it's supposed to say for all of these people what does their resting blood pressure look like as a snapshot.

I think the same with some of the other measures, such as the glycosylated hemoglobin, being an average snapshot of what probably is going on, and then not trying to characterize for individuals but really using the population sample as representing groups of people where on the whole we think for the men we've captured an average picture of what men look like with low SES.

I think this gets to John's discussion earlier about meeting all these different levels of analysis that these population studies are never going to be the ones that really look at those time courses very much. That's where you're going to need to get these smaller samples where you can afford, because you have fewer people, to do the much more detailed monitoring over time for multiple days or that kind of thing.

So I think to that question, we're not dealing with those different time courses. We're trying to get sort of a snapshot, and then somebody else needs to be looking at the subset needs to be looking at the time courses so you can integrate the time course with the snapshot.

MAHAY: Take about three more questions.

LAUDERDALE: I wanted to ask on the same basis why you include using spirometry?

SASTRY: To look at asthma in kids.

LAUDERDALE: I understand that, but people with asthma have normal lung function if you get them at any old day, so what are you going to gain?

SASTRY: That they do their own lung function test regularly?

HALTER: If they're having say an acute event, their spirometry will look much different.

LAUDERDALE: That's the definition of asthma, the normal between time; so if you just walk in the house, they're going to have normal spirometry.

SASTRY: No, I don't think that's the case. I think there are subtle differences in that profile. That's the difference from doing peak expiratory flow to doing spirometric measure is that you can get some sense of how they perform in these different aspects.

I'm not an expert in this. I'm not trying to argue too far, but that's what the epidemiologists who we are working with say; and that's actually the standard that's being done. It's not asthma actually. It's broader than that. I think there are well-known standards for how kids' lung development looks by age and by sex and that you can really look at whether somebody has diminished lung function based on their exposure to pollution and so forth.

There's an interesting set of findings coming out of California from studies in Sacramento and certainly in Los Angeles that have tracked kids over time and looked how exposure to pollution, other factors, have shaped lung development.

LAUDERDALE: I suspect pollution issue is (inaudible) asthma.

SASTRY: I think they're actually related pretty closely, but I can't tell you the details.

McDADE: Since you're apparently stuck with the phlebotomists in California, have you thought about actually collecting venipuncture blood from at least a subsample? That might address a lot of issues you have here. You could use that to validate some of the blood spots, opens up additional markers in the future. You might be able to use that as an opportunity actually.

SASTRY: Yes, yes. It just adds in another layer of maintaining the cold chain.

McDADE: It does. The cold chain is an issue, but it would be a marginal additional cost.

SASTRY: Part of our hope was that the finger prick samples would be easier; people would be more willing to do it. If you're doing it for assessing some of the validity of these things, it would have to be --

McDADE: On a subsample.

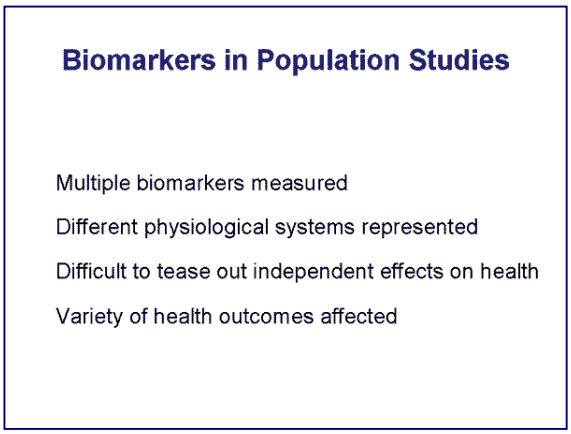
MAHAY: Thank you so much.

UCLA, CARDIA & MacArthur Population-based Studies

Arun Karlamangla, M.D.

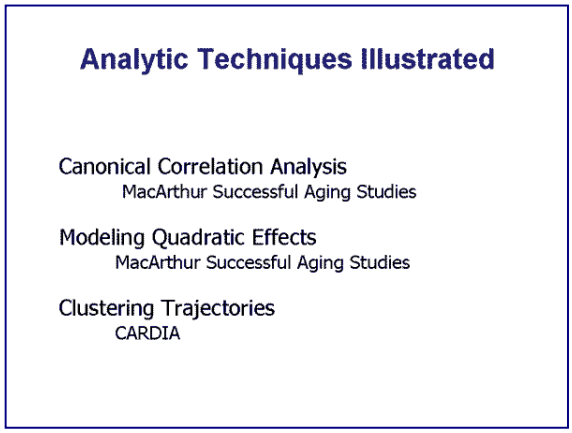
I was asked to speak about strategies that we've used to analyze multiple biomarker data from population-based studies. I don't intend this to be a comprehensive overview of all possible analytic strategies one could use with large population-based studies, but just an illustration of different techniques that we have used.

So what's typical about these studies? They are large studies; multiple biomarkers are measured; you have multiple hypotheses regarding psychosocial predictors for each one of these biomarkers, and often there are also multiple health outcomes measured- death, disability, fractures, hospitalizations, et cetera, et cetera. Now, the biomarkers represent different physiological systems, and even in large studies it's often difficult to tease out the independent effects of different biomarkers. A single biomarker, let's say C-reactive protein, might affect a multitude of outcomes and not just one outcome.



Also, especially in older adults, the people that I'm most interested in, individual biomarkers may have smaller associations with outcomes, but if you combine all the different biomarkers, the combined effect of small elevations in multiple different physiological systems can really be substantial and lead to bad health outcomes. So the question from an analytical perspective is how can we combine multiple biomarker measurements to best capture their combined effect on health.

Now, there might be other reasons which might motivate you to combine multiple biomarkers into a single score. One possible reason might be, if you didn't, if you instead tested multiple hypotheses with 10 different psychosocial predictors with 20 biomarkers, you have a possible 200 different hypotheses, and you very quickly run in to multiple hypothesis testing problems. If, on the other hand, you have a semi-intelligent way of combining the biomarkers into a single score, one predictor versus one biomarker score, you've reduced the number of hypotheses tested by an order of magnitude.



So the analytical techniques that I've chosen to illustrate here today are three that we used in two different population studies - the MacArthur Successful Aging Study and CARDIA. CARDIA was a study of the growth of risk factors in young adults. The techniques that I'll address are canonical correlation analyses, models with quadratic effects, and empirical clustering of growth trajectories. In CARDIA they measured these cardiovascular risk factors over multiple visits, five, now six visits. So our objective was to look at how these biomarkers changed over time as these young adults grew into their middle age, and we wanted to identify

Canonical Correlation Example

Cohort: MacArthur Successful Aging Studies

Background: A simple count of the number of biomarkers in the worst quartile has good predictive ability for future functional decline

Question: Can we improve the prediction ability by more optimally combining biomarkers?

MacArthur Aging Study

Study Sample:

70-79 year old community-dwelling

N=1189; top 1/3 of ~4000 screened for cognitive and physical functioning

Baseline measurements in 1988

Follow up measurements in 1991 and 1995

MacArthur Biomarker Measurements

Blood Pressure - Systolic and Diastolic
Glycosylated Hemoglobin
Waist-hip Circumference Ratio
HDL Cholesterol
Total Cholesterol
Overnight 12 hour Urine Epinephrine
Overnight 12 hour Urine Norepinephrine
Overnight 12 hour Urine Cortisol
Serum DHEA-S

clusters of growth trajectories in order to then look at how membership in different clusters varied with demographics and lifestyle and other factors.

The canonical correlation example is taken from our analysis of the MacArthur dataset. The background here is that previous research had shown that a simple count of the number of biomarkers in the worst quartile had good predictive ability for multiple different health outcomes. This count of the number of risk factors in the top quartile was called allostatic load in Teresa's paper.

The questions that we raised were: Can we improve on the prediction ability of these biomarkers by including more information from each, not just looking at whether or not you're in the worst quartile? Shouldn't the actual magnitude of the biomarkers have some predictive role? Shouldn't different biomarkers be weighted differently?

Why give them all the same weight? A little introduction to the MacArthur Aging Study. It was drawn from the population from community-dwelling older adults in their 70's. They were chosen to be high-functioning; they were in the top 1/3 with respect to performance in both physical and cognitive tests. Baseline measurements of these folks were made in 1988. The first follow-up was 2-1/2 years later, in '91, and the final follow-up in '95, giving you a total of 7 years of follow-up.

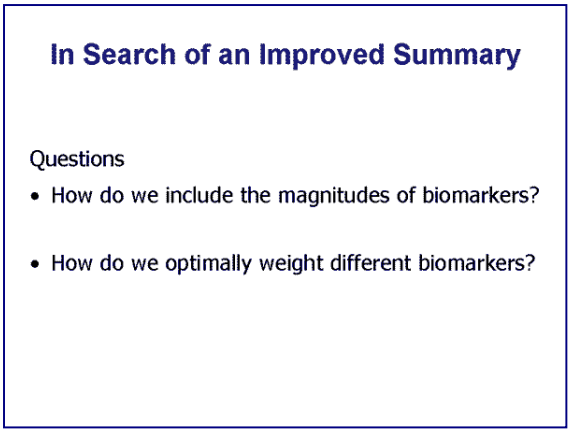
The usual cardiovascular risk factors were measured. In addition, there were measurements of what Maxine called preclinical or subclinical factors - neuroendocrine hormones, epinephrine, norepinephrine, cortisol, and DHEA.

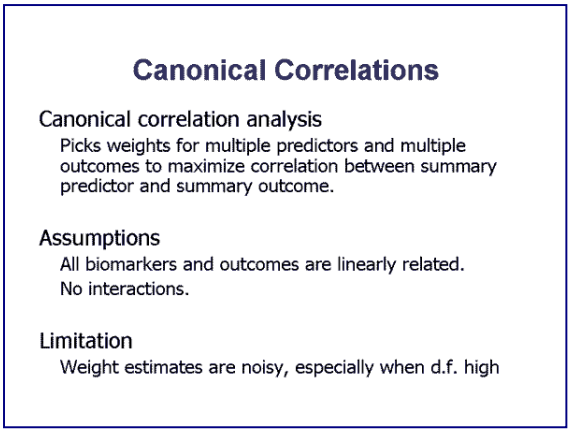
There were lots of outcomes looked at. We were interested in functional decline in these adults who were initially high-functioning. So we looked at declines in performance scores on a battery of 5 physical performance tests and 5 cognitive performance tests, and we had 2 different time periods over which we had these outcomes - between baseline and the first follow-up and between first follow-up and second follow-up. So just looking at these outcomes, we have 10 tests, 2 decline periods, and that's 20 decline-scores.

In searching for an improved summary of biomarkers, how can we include the magnitude of the biomarkers and how should we weigh the different biomarkers? Those were the two questions.

We chose canonical correlation analysis to address them. Briefly, about canonical correlation analysis: Given a set of predictors and a set of outcomes, it gives you weights for the predictors and weights for the outcomes, so that the weighted sum of the predictors is maximally correlated with the weighted sum of the

outcomes. It combines both sets of variables into one variable each, maximally correlated with each other, and that maximal correlation is called the canonical correlation. The best weights are called the canonical weights. The assumptions, of course, are that the predictors and the outcomes are linearly related and that there are no interactions. A big limitation with this approach, and you will see this described in statistics texts, is that the canonical weights you get are very noisy; that is, they are very sensitive to different data points. If you drop one or two data points, you get a big variability in the canonical weights. So how might we overcome that? One possible approach, the approach that we adopted, was to look at the bootstrap distribution of the canonical weights. By the way, I didn't mention at the very beginning the people I worked with closely in all of these analyses, Teresa Seeman and Burt Singer. Burt Singer was my statistical mentor for all of this. He is on the faculty at Princeton. So he suggested that bootstrapping seemed to be -- yes?





HALTER: How do you deal with the assumption about no correlation among the variables?

KARLAMANGLA: No, no, that wasn't the assumption.

HALTER: Oh, I'm sorry.

KARLAMANGLA: The assumption was that the --

HALTER: No interactions.

KARLAMANGLA: No interactions, correct.

HALTER: So the fact that there are intercorrelations among physical performance measures is not an interaction?

KARLAMANGLA: No, no. Interaction would mean, let's say, that smoking affects heart disease only in people with hypertension and smoking does not have an effect on heart disease in people who do not have hypertension. That's an interaction, not simply that smoking is correlated with hypertension. "Effect modification" is the other term that's used for interaction.

Variables can be correlated and they are often correlated, which is why we are looking for combinations --

HOUSE: If you preidentify interactions, you can build in sort of interaction terms that might be viewed (inaudible).

KARLAMANGLA: I suppose we could, yes. It's like, say, if you knew a priori that one of the biomarkers also had a quadratic effect, you might be able to put a quadratic term in there. There's a problem with doing that, and the problem is that the standard error estimates and the p-value estimates that you get assumes something about the distributions of these terms, and once you put interaction terms or squared terms, the distributions of the terms are no longer bell-shaped and the p-values you get aren't reliable. So you have to get around it, and what I suggest in the next slide here might be one way of getting around that. So yes, we could probably get around the problem with interactions with fancier approaches.

Canonical Correlation Analysis:
Overcoming the instability in estimates

Bootstrap distributions can be used to assess estimation error in canonical weights .

Large error, relative to the estimate itself, reflects a marginal contribution of the biomarker / outcome to the association

Stepwise backward elimination of predictors and outcomes with the largest estimation errors will lead to a pruned model with stable weight estimates.

Canonical Correlation Example

MacArthur. Predicting Functional Decline

Predictors: 10 biomarkers measured in 1988
Outcomes: Declines in 10 performance scores over 2 periods: 1988-91 and 1991-95

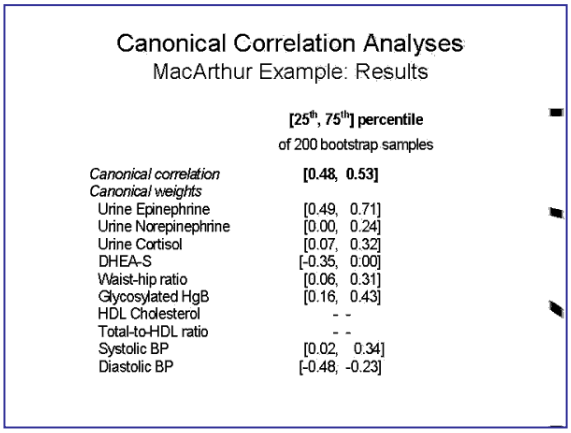
Bootstrap distributions of canonical weights: Based on 200 bootstrap samples.

Stepwise backward elimination: Predictors / outcomes standardized. Those with the widest bootstrap distributions (of canonical weights) dropped stepwise till no interquartile range crosses zero.

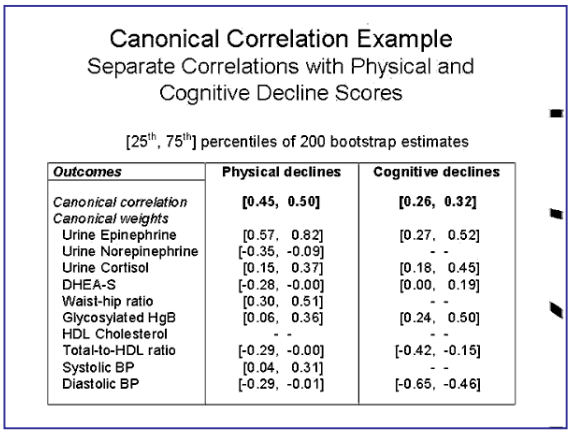
One of the major limitations, as I said, is the variability in the canonical weight estimates, so we decided to use bootstrapping. If you look at the bootstrap distribution of the canonical weight, that gives you an estimate of the standard error. The standard deviation of the bootstrap distribution of the canonical weight is the standard error in the canonical weight. A particular predictor or outcome that has a large standard error is not contributing significantly to the relationship between biomarkers and outcomes and can probably be dropped. So we used stepwise backward elimination, dropping variables with large standard deviations in the bootstrap distribution, and then you get a pruned model, and in the final model the weights are more stable.

In the MacArthur dataset, this is what we have: 10 biomarkers measured in 1988; declines in 10 different performance scores over 2 periods. We're looking for weights on both sides. We used 200 bootstrap samples to assess the variability in the weights. The predictors and outcomes were standardized first, and then those with the widest bootstrap distributions of the canonical weights were dropped one by one in a stepwise fashion until the middle half of the distributions of the weights did not cross zero. At that point, we stopped the backward elimination. We terminated the backward elimination when the interquartile range for the bootstrap distributions did not cross zero.

These are the canonical weights we obtained. What you have here are the interquartile ranges for the different canonical weights. The very first row gives you the interquartile range for the canonical correlation. You get a correlation of about 0.5, which is significantly higher than the correlations obtained with the simple count of risk factors, which were in the 0.1, 0.2 range. And here are the weights. Interestingly, lipids dropped out. As one of the other speakers pointed out earlier, in an older population, lipid levels don't seem to matter.



Because the outcomes we combined were mixed, both physical and cognitive, we decided to also do the analyses separately for physical and cognitive declines, and the weights are similar, and now the total-to-HDL cholesterol ratio dropped back in and a few others dropped out for cognitive declines. Notice that the correlation with cognition wasn't as good as the correlation with physical function decline.



HOUSE: What's the measure that's on the so-called dependent variable side here, the declines? You're using like a difference between measure?

KARLAMANGLA: Yes. Do you want to see the slide?

We had five different tests of physical performance, five different tests of cognitive performance. We had declines in the scores. The scores were standardized, zero to 1, and then we had declines in these scores between '88 and '91 and between '91 and '95. So once you do the canonical correlation analysis, what you get is a weighted sum of all of these decline scores and a weighted sum of all the biomarkers, and the correlation between them was .5

KURINA: Do all of the outcomes always stay in the models?

KARLAMANGLA: No. Some of the outcomes dropped out too.

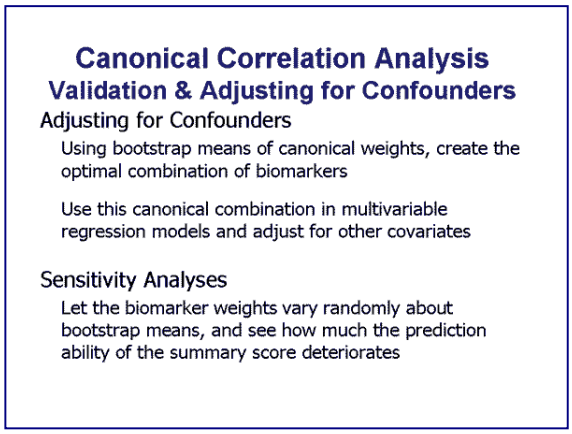
KURINA: Did some of those outcomes drop out in those?

KARLAMANGLA: Yes, some outcomes dropped out too. I didn't put that in here.

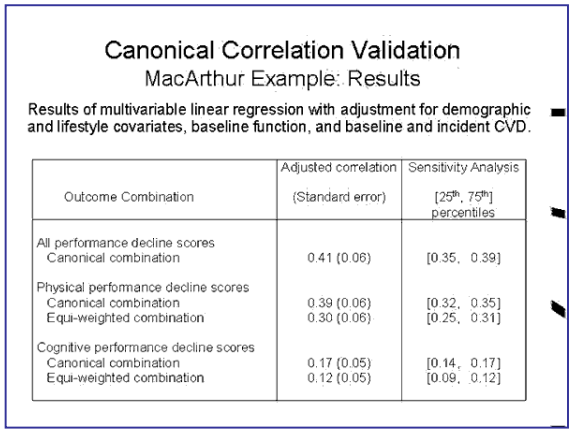
KURINA: I was actually interested in the question though. I'm sorry.

KARLAMANGLA: I can send you that information later.

So some of the outcomes dropped out, too, which means that their weights were very variable and they weren't contributing to the association as well.



Okay. Now, when you build a model like this using the data, you don't know if the answer that you get is peculiar to this particular data because the same data was used to build the model, so you have to do some kind of validation. Another thing, of course, is adjusting for confounders, such as demographics and lifestyle and so on. So we used the means of these bootstrap distributions of the canonical weights as the weights and created a combined score on the biomarker side, a combined canonical score on the outcome side, and then used regression models to predict the canonical outcome as a function of the canonical biomarker score and with other confounders in the model. Then we did a sensitivity analysis where we let the weights vary randomly around the optimal weight to see how much the association depended on the choice of weights.



The first column here gives you how the correlation changed after we adjusted for confounders. We adjusted for the usual confounders, age, gender, ethnicity and all of that, and cardiovascular disease as well. The correlation dropped from .5 to .4 after adjusting, and then we did sensitivity analysis and it went down further. Well, you might wonder when I did the sensitivity analysis why did the range now go away from .4. Remember, .41 is the best you can do. If I vary the weights around that best point, the correlation is going to drop. So the distribution actually drops from .41, and so you get smaller correlations when I let the weights vary around. And then because some outcomes had been dropped out, we also looked at simply taking an equi-weighted combination of all the outcomes. Because the outcomes were chosen by the people who designed the study, they may have a certain value, and they had used equi-weighted combinations in the past. So instead of using the canonical combination of outcomes, we used the equi-weighted combination of outcomes and noticed that the correlation drops a little bit.

WEINSTEIN: Did you find that everybody declined in their cognitive performance, and is that because you had a high-functioning group to begin with, so that all that could happen is that they would go down?

KARLAMANGLA: Good question. We had a high-functioning group. High-functioning was defined in terms of whether or not they had disabilities by ADL's, the activities of daily living, so rather gross definitions of high-functioning, and whether or not they scored 6 or higher

on the Short Portable Mental Status Questionnaire. So when you look at these rather gross definitions of disability and you drop out the people who are poor-functioning, you get what's called a high-functioning group, but then you go and look at these finer tests of performance and you get a nice bell-shaped distribution for all the scores, which means there is room for people to go both up or down, and, in fact, some people did improve and the majority went down. In the first follow-up period, the mean change was close to zero, but in the second follow-up period, the mean change was negative. So the majority declined, but not everyone declined.

LINDAU: How long is the follow-up period?

KARLAMANGLA: The first was 2-1/2 years and the second was 4-1/2, a total of 7, and these are people who were in their 70's, 70 to 79.

WEINSTEIN: So when you say "cognitive performance decline," it's "decline" in sort of quotation marks because some of them may actually have improved.

KARLAMANGLA: When I say "decline" here, I mean the change where change is defined as baseline minus final.

WEINSTEIN: So some could be positive and some negative?

KARLAMANGLA: Some could be negative, correct.

HALTER: What about directionality of the markers?

KARLAMANGLA: That was in the previous correlation table.

HALTER: So you have, I believe it was positive for epinephrine and negative for norepinephrine, or the other way around? I didn't understand that.

KARLAMANGLA: I don't want to go back, but what we found was what you would expect. The lipids dropped out. DHEA --

HALTER: Norepinephrine and epinephrine.

KARLAMANGLA: They were both positive, if I can remember right.

(Multiple voices.)

KARLAMANGLA: Diastolic blood pressure and DHEA-sulfate were negative, which meant that a high value of DHEA-sulfate is protective and a high value of --

(Technical difficulties.)

LINDAU: We're having technical difficulties, so the degree to which we can use the microphones would be very, very helpful.

KARLAMANGLA: Okay. Quickly, in this you'll notice that just the blood pressure and the -- just a second. Okay. In this, diastolic blood pressure and DHEA were the only ones negative, and this one -- let's see.

HALTER: The epinephrine is positive and norepinephrine is negative.

KARLAMANGLA: Yes. So what that means is, after you adjust for epinephrine, -- there's a lot of correlation between epinephrine and norepinephrine. They are both catecholamines.

HALTER: Right.

KARLAMANGLA: The independent association of this in this case turned out to be negative, because after I adjusted for epinephrine, norepinephrine was negatively associated.

ADAM: So simple association is positive?

KARLAMANGLA: Simple association probably is positive, yes.

HALTER: So you're saying adjusting for epinephrine, norepinephrine actually is associated with physical improvement?

KARLAMANGLA: Yes.

HALTER: Is that how you interpret it?

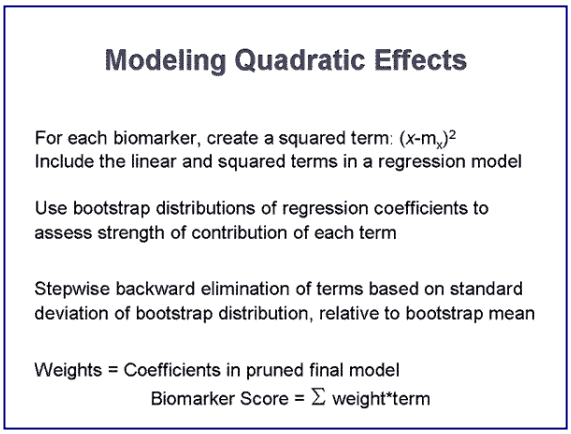
KARLAMANGLA: Yes, yes.

HALTER: This is kind of weird.

KARLAMANGLA: Yes, I agree.

(Multiple voices.)

KARLAMANGLA: But anything else is as you would expect. Diastolic blood pressure, DHEA-sulfate, and total-to-HDL ratio, which, as someone had mentioned earlier, can be protective; high cholesterol can be protective.



All right. So the next technique I was going to illustrate was using quadratic effects. In the canonical approach, as I said, I assumed that everything was linearly related. What motivated us to look at quadratic effects was the observation that both low and high values of some biomarkers might contribute to risk. Blood pressure is a classic example. High blood pressure, hypertension, is a cardiovascular risk. Low blood pressure, at least in older adults, is often associated with passing out, with syncope, falling, fractures and all kinds of things. The same is true for measures of abdominal obesity. People who have a greater waist circumference often have more cardiovascular and people who have low waist circumference probably have cancer or something and are declining. So how do you try and model both? What we decided to do was create a squared term, where for each of the variables, we subtracted out the sample mean and squared the deviation from the sample mean to create a

squared term, and then put both the main term and the squared term in the regression model. Then once you now have squared terms, the problem that I spoke about originally comes back here again, and that is that the p-values and standard errors are no long reliable since these squared terms are not distributed nicely on a bell-shaped curve. So I pull the same trick out of the sleeve - bootstrapping. You look at the bootstrap distributions of the regression coefficients, and from that distribution get the standard error. The standard deviation of the bootstrap distribution of the regression coefficient is the standard error for that regression coefficient. So we used that information to do backward elimination, and then after we eliminated, in the final model, we just combined each term with its weight to create a biomarker score.

So this example, again, 10 biomarkers, as before, measured in 1988. We looked at all-cause mortality over the total 7-year follow-up. We used logistic regression with both linear terms and squared terms. Using 200 bootstrap samples at each step, we eliminated terms, and we terminated when the standard error was less than the parameter estimate.

Quadratic Model Example

MacArthur. Predicting All-Cause Mortality

Predictors: 10 biomarkers measured in 1988
Outcomes: Total mortality over 7 years (till 1995)

Multivariable Model: Logistic regression with squared terms

Bootstrap distributions of regression coefficients: Based on 200 bootstrap samples.

Stepwise backward elimination:
Until bootstrap standard deviation <= |bootstrap mean|

So these are the regression coefficients that we were left with after all of this.

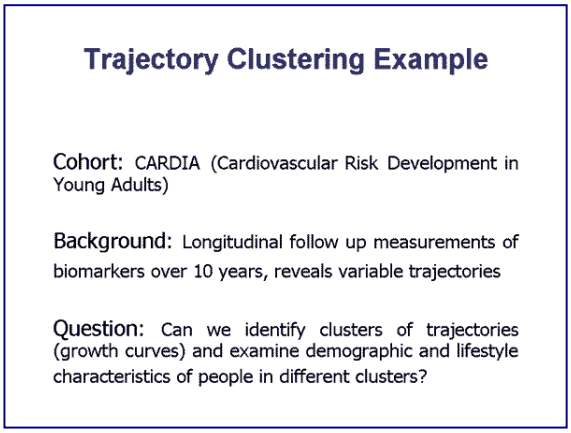
ED LAUMANN: I'm just wondering, how does reliability and validity of these total measures play into this process?

KARLAMANGLA: The measurement error is what you're talking about?

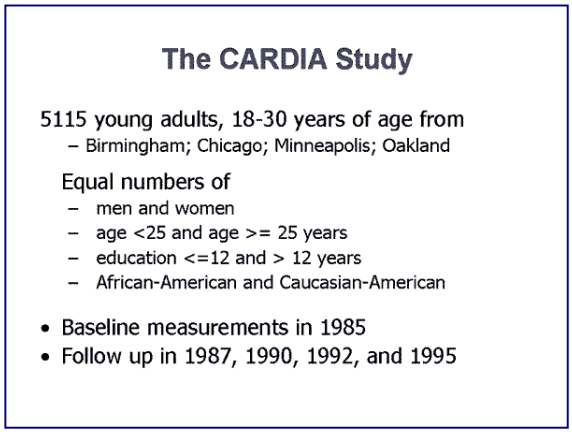
ED LAUMANN: Well, it varies across the measures, so that some measures, we were looking at lower urinary tract syndrome (inaudible).

KARLAMANGLA: That's a very good point. I hand-waved a bit earlier when I said that if the standard error is large, it means the variable makes a small contribution to the association with the outcome. That's only partly true, assuming that all the variables are measured with equal amount of measurement error, because measurement error also contributes to the error in estimating the regression coefficient or the canonical weight for that particular variable. So our backward elimination process eliminates variables that are either measured poorly or make a small contribution or do both badly.

All right. So this is what we were left with. The actual numbers aren't important. I'm illustrating the strategy here. But I just wanted to show you that a lot of the quadratic terms stayed back in there, not surprising considering this is an older adult cohort. For blood pressure, only the quadratic term stayed in, the main term did not. Waist-hip ratio - only the main term stayed. For glycosylated hemoglobin, there's a main term and a quadratic term. Lipids now stayed back in there as main terms but with a negative coefficient, meaning that a higher value of the lipids was actually protective against mortality.



And then we used this combined biomarker score, and this is how it predicts mortality. These are the deciles of the biomarker score, and that's 7-year mortality on the Y-axis. As before, we have to worry about the fact that the model was created from the same dataset in which it is being tested, and we need to adjust for covariates. Just as before, we can use sensitivity analysis, we can use bootstrapping, and all of that. In this example, we adjusted for age and gender, and we found that age, gender, and the biomarker score each independently predicted mortality. At this point I looked for interaction between age, gender, and the biomarker score, and we didn't find any.



The third technique I want to illustrate is the analysis we did in CARDIA where we looked for clusters of trajectories of biomarkers. This is a cohort of young adults from four different communities in the U.S. Longitudinal follow-up measurements were made of these biomarkers over 10 years, and we found that different people had different kinds of trajectories. Some increased, some decreased, and some stayed the same. We wanted to identify these clusters and then correlate the membership in the cluster with socioeconomic status, which is the question we were addressing in this case.

A bit about the CARDIA study: More than 5,000 young adults, young defined as between 18 and 30 years of age at baseline, but equal numbers of men and women, people below 25 and above 25 years of age, education less than high school and more, African-American and Caucasian-American. So this is a nice mix. Baseline measurements in '85, and then did measurements about every two years after that until '95.

LINDAU: Is this a clinic-based sample or population-based?

KARLAMANGLA: Population-based. They are from four sites - Birmingham, Alabama; Oakland, California; Chicago; and Minneapolis, Minnesota.

LINDAU: Thank you.

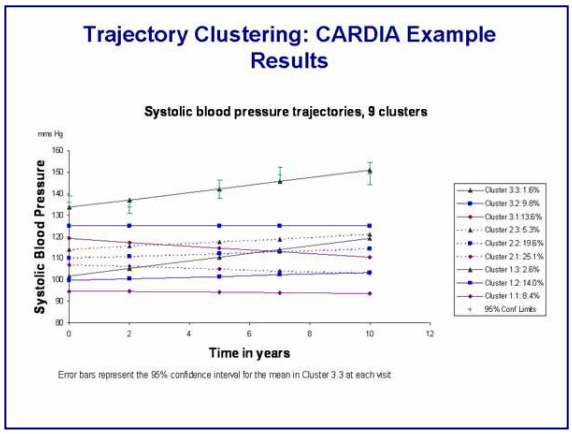
KARLAMANGLA: This is a cardiovascular study funded by the NHLBI.

Socioeconomic status measurements were pretty good. They had education of the parents, education of the participants themselves, and they also had a question about financial hardship in paying for basics. Now, the participant education analysis we did a bit differently because many of these folks were still getting their education at baseline, and so I'm not presenting those results, just for the other two.

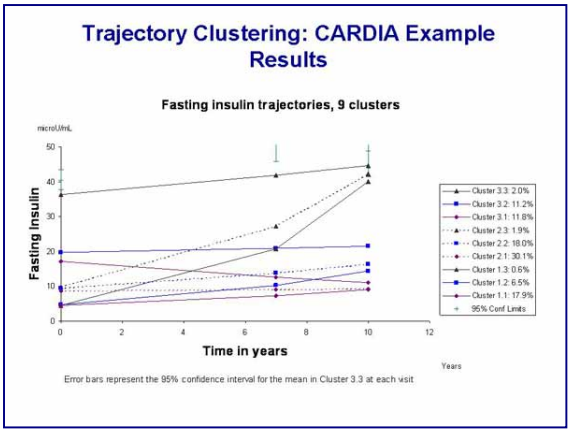
To identify clusters of trajectories we used a semiparametric mixture model, which was developed by Nagin, and the software was written at CMU by Bobby Jones and Nagin. It's available on the web. It's for SAS and is called PROC TRAJ. It finds population subgroups with similar trajectories and then fits a single trajectory to the cluster. So it divides the population into multiple clusters, depending on similarity, based on homogeneity of trajectories, and then gives you a common representative trajectory for each cluster parametrically. The breaking up into clusters is why it is a mixture model and each cluster's representative trajectory is parametric; hence the semi-parametric mixture.

Unfortunately, the clustering of the trajectories depends on where you start from. So at baseline, if your biomarker values are quite different, you're automatically put into two different clusters even if the shape of growth is the same over follow up, because that baseline difference tends to dominate. So what we decided to do to overcome that problem, because we're interested really in how the biomarkers change and not just clustering by baseline value, we first stratified the study sample by the baseline measurement. We divided the sample into three groups based on the baseline value of the biomarker for each biomarker - the top quartile, the middle two quartiles, and the bottom quartile, and then we did the clustering within these three groups separately.

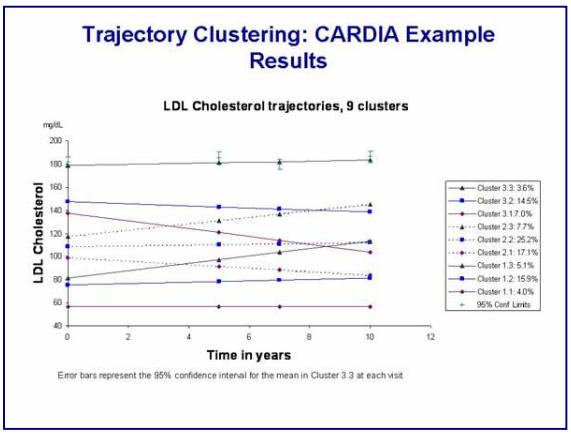
Here is an example of the clustered trajectories of one of the biomarkers. You can identify three strata here. The black, blue, and red here is for the top stratum - the top quartile, the middle two quartiles - the middle stratum, and the bottom stratum – the bottom quartile, respectively. So within the stratum of people who started over here, some went up, some stayed the same, and some went down, and the legend gives you what percentages of people did each. The same thing happens in each of these three strata. The software gives you these representative trajectories and tells you who is in which trajectory.



This is a similar kind of thing for fasting insulin. There was exponential growth in insulin for some people who started low. Others grew like that.



This is a similar thing for LDL cholesterol. We have similar clustering for every one of the biomarkers.



So what did we do after we did this clustering? For each of the 7 biomarkers, we then assigned a score. Wherever you were at baseline, if you then increased, if your biomarker increased, you got a score of plus 1. If you came down, since we believe that coming down from a high value is good for you if you started from up on the top, that group of people who started on the top and then declined got a score of minus 1. So other people who start in the middle or were in the lower end to begin with and then went down normally didn't go down that much anyway, so they didn't get the benefit of a minus 1. So the plus 1 was everybody who grew and minus 1 for those who declined from the top.

Combining Cluster Membership Across Biomarkers

For each of the 7 biomarkers, assign

- - 1 point to the declining cluster in the top stratum
- +1 point to the top-growth cluster in each stratum
- 0 points to all other clusters

Combined Risk Accumulation Score
= Sum of scores over all 7 risk factors.

CACIOPPO: I'm just curious as to why you didn't use the same logic, since you showed evidence that the quadratic being very low is also harmful, so those who were in the bottom stratum who showed an increase didn't show an improvement rather than showing that as a decline. Do you see what I'm saying? If you have a quadratic and you start very low, moving up should be healthier, salubrious, rather than damaging by your earlier logic, correct?

KARLAMANGLA: Oh, the earlier logic was that if you're in the middle somewhere, that's good, and then if you go out from it in either direction, that's bad.

CACIOPPO: Right. So if you start out very high and you're moving down, you're saying that's good?

KARLAMANGLA: That's good.

CACIOPPO: But if you're starting very low and you're moving up, that's not good?

KARLAMANGLA: It depends on what you mean by very low. See, this is quartiles. First of all, this is a population of young adults where the range at the very beginning is not too high. If you look at the numbers there, look at what is low here. A low cholesterol of 80, a low LDL of 60 or 80 is not a bad thing, and the same thing with blood pressure and so on. These people were not hanging around with a systolic blood pressure of 50 or 60. They were still good numbers, 80 or 90. So yes, I threw away the idea of the quadratic at this point.

All right. So what we did here was plus 1 for growth, minus 1 for a decline if you decline from the top, and then we added these scores over all the 7 biomarkers to get a comprehensive risk accumulation score. Then once we had a risk score, a single number summarizing growth in biomarkers over the 10 years, growth in all the biomarkers all together summarized in one number, risk accumulation score, we then looked at its associations with the SES markers, parental education and financial hardship, and these are the correlations we found after adjusting for – I forget what they were adjusting for. I think adjusting for age, lifestyle and a few other demographic variables.

The correlations here are positive correlations. The reference group here are people who had high school education. So less than high school education, greater risk, and college graduate, you get reduction in risk. The trends were not significant in every one of the groups. They were most significant in the white women and least significant in black men. The same thing with financial hardship; those who reported financial hardship tended to have a higher risk accumulation score. Again, it was most significant in black women here.

CARDIA Example: SES Associations

SES variable	Adjusted Effect on Risk Accumulation Score (Standard Error)			
	Black men	White men	Black women	White women
Parent Education				
< High school	+ 0.09 (0.16)	+ 0.18 (0.25)	+ 0.10 (0.12)	- 0.03 (0.15)
1-3 years college	+ 0.16 (0.16)	- 0.03 (0.16)	+ 0.02 (0.11)	- 0.00 (0.09)
College graduate	- 0.03 (0.15)	- 0.23 (0.11)	- 0.10 (0.11)	- 0.21 (0.07)
	(NS)	(p=.05)	(NS)	(p=.003)
Financial Hardship	+ 0.22 (0.12)	+ 0.19 (0.10)	+ 0.22 (0.08)	+ 0.16 (0.07)
	(p=.05)	(p=.1)	(p=.007)	(p=.02)

JORDAN: Since your study went over a 10-year period and you had multiple sites, did you require that all of your biomarkers be run on the same clinical platform at the sites and the same kit, because there can be subtle variation in an answer with different platforms, different kits. Did you control for that?

KARLAMANGLA: I actually don't know the answer to that.

SEEMAN: As with many of the multisite studies, CARDIA has central labs that run them all. The different sites all ship their samples to a central lab. So they are all done through the same lab.

JORDAN: And that platform hasn't changed over time?

SEEMAN: No.

KARLAMANGLA: I just want to continue. There are a variety of ways one could combine biomarkers, provided you want to combine biomarkers. In clinical practice, and I'm coming from a clinical practice background, I find utility in doing this. We already use combinations, such as the Framingham risk score, to predict who is going to get cardiovascular disease. We use the metabolic syndromes. So these are similar ways of doing it. The problem

with the Framingham risk score, for instance, in older adults is that it just does not work. The original Framingham cohort didn't have anyone over 70, and those results aren't easily translatable to older adults. So there are other ways of doing it. Here are three methods that we used.

I welcome comments.

(Applause)

FENDRICH: There are some Bayesian statisticians who have used methods for combining multiple diagnostic outcomes. Have you looked at any of that or has anybody on your team?

KARLAMANGLA: No, I haven't. With the Bayesian approach, you always have the problem of having to have an a priori hypothesis or an a priori estimate of what the associations are. No, I haven't.

WEIR: I have a question about the horizontal axis on your last thing. It's elapsed time in the panel that you're measuring these trajectories over.

KARLAMANGLA: Right.

WEIR: Many of us would rather think of that in terms of age, for example. I don't know if there is any work being done on how to model those trajectories as functions of age rather than time elapsed. Secondly, within the real world where we don't get people to do things on the exact day we wish they would, those intervals aren't always of equal length, and I don't know whether that affects how those models work.

KARLAMANGLA: Good point. The trajectory clustering program doesn't care if it's age or time. You could do it either way. The program does, of course, assume that the measurements were all equally spaced. But there are things we could use. One could use things like a random effects model which would allow you to take into consideration that different people were sampled at different intervals.

GAVRILOV: Did you try to analyze data for males and females separately?

KARLAMANGLA: That's a good question. In this I assumed that the optimal combination of the different risk factors - the different biomarkers would be the same in both men and women, which is not really true. As was pointed out in the very first talk today, you want to combine analyses at different levels and you want to bring gender into the picture or some other social, demographic characteristic into the picture. One could do that in all of this, provided there is no interaction. What you're asking about is the interaction between the different biomarkers and gender. Once you bring in interactions, there are so many different interactions that are possible, you have to have some a priori hypothesis and then stratify by gender. Suppose an a priori hypothesis is that they are different by gender or they are different by ethnicity; you would have to do the model building separately within each group.

GAVRILOV: But you can split the sample in two parts, males and females, and analyze just to see whether the results are similar or not. It's a simple way to do it.

KARLAMANGLA: True, break it up into groups and do the analysis in each stratum.

Proper Use vs. Misuse of Biomarker Data in Social Science Research: The Case of Cortisol

Emma Adam

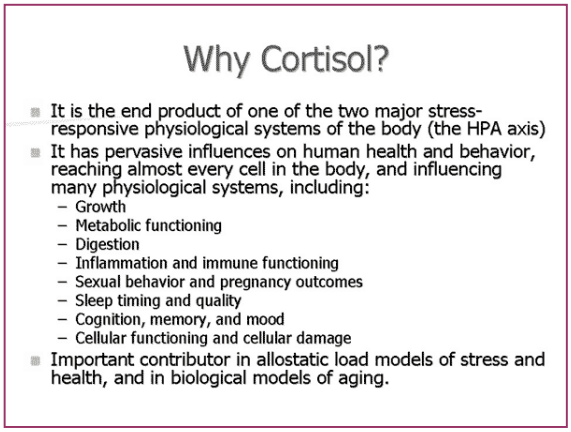
I'm happy to talk to you today about cortisol. Thom originally named the symposium, and I know he was joking, but he originally named it "Biomarkers Gone Bad," and I've renamed it "Proper Use vs. Misuse of Biomarker Data In Social Science Research: The Case of Cortisol," because I want to make the point that you should never blame the biomarker for our failure to understand it. It's just an innocent biomarker, and we scientists are the problem, not the biomarker.

I'd like to start out by thanking the various funders of my research. I'm not going to be talking to you much about my substantive research. I'm mainly going to be talking methodologically today, but all of my research is focused on understanding how people's everyday life experiences get under the skin to influence the activity of the HPA axis, and I tend to study those processes for both parents and in children within the context of the family environment.

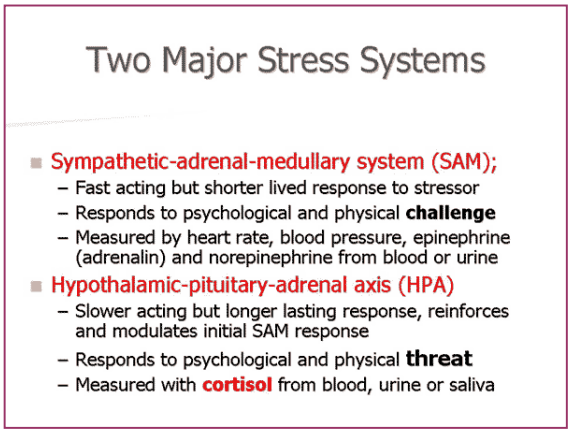
I'm going to cover quite a few things in the 45 minutes that I have, so I'll probably speak rapidly and cause stress in our transcriber. I'm going to first of all cover "why are we interested in studying cortisol in the first place?" and then talk about the importance of proper measurement approaches and the importance of gathering multiple samples, at least in the case of cortisol. I'll define a variety of different ways you can look at cortisol data, each of which has a different meaning and interpretation, I call these different "cortisol parameters". Then I'll consider some of the multiple contributors to different cortisol parameters, and introduce two approaches that attempt to isolate the meaningful variation in cortisol and throw away some of the junk variation. There are various statistical approaches that you can use to do that in cortisol, and I think that some of those will transfer over to thinking about other biomarkers as well. If you've measured your biomarker better and if you're putting only the trait variation associated with that biomarker into the model, it will result in a better, more predictive model. Then if I have time, I'll talk a little bit about compliance concerns regarding ambulatory cortisol sampling and some of the solutions that we have come up with for that, and finally speak briefly to the importance of theory in conducting cortisol and other biomarker research.

A grant reviewer once said to me in their review, "Cortisol will break your heart." That was the most concise review I've ever received. My response to that reviewer was, "It will only break my heart if I don't respect it, understand it, and measure it properly." I did go on to get that grant. The reviewer's second statement in the review was, "Best of luck. I hope you succeed where others have failed." I do think that there is substantial meaningful variation associated with cortisol activity that we can get at and that is important to get at. The point I want to make here is that biomarkers are complicated. They are influenced by and also influence many aspects of the internal and external environment. We're not measuring them because they're easy to measure. We're measuring them because they represent an important causal variable in models of the influence of societal factors on health and development. Now, I do understand when you get into the realm of population-based studies there are important reasons why ease of measurement is important, but I don't think we should lose sight of the fact that the variables we choose are the variables that we most need for our models and not the variables that are the easiest to capture.

So let me say a little bit about why we do want cortisol to be one of our variables amongst all the very other important physiological parameters we're looking at. First of all, as we probably all know, cortisol is the end product of one of the two major stress-responsive physiological systems of the body, the HPA axis. All the clinical literature suggests that cortisol has pervasive influences on human health and behavior. It reaches almost every cell in the human body and influences many other physiological systems, including the growth axis, metabolic functioning, digestion, inflammation, immune functioning, sexual behavior, pregnancy physiology, timing, and outcomes, sleep timing and quality, cognition, memory, mood, cellular functioning, and even cellular damage.



In our inaugural conference of Cells to Society at Northwestern we had the pleasure of having Rick Morimoto there, who studies the role of protein folding and misfolding in disease processes. One question asked of him was, "how is this influenced by the external world?" He noted that there are multiple pathways by which broad social events as well as physical environmental events can have an impact, and one pathway was through the effects of the HPA axis and cortisol on cellular processes. Cortisol is thought to be an important contributor to allostatic load - it's a primary mediator in allostatic load models of stress and health. It's also an important player in many biological models of aging. So from a theoretical standpoint, it's important.

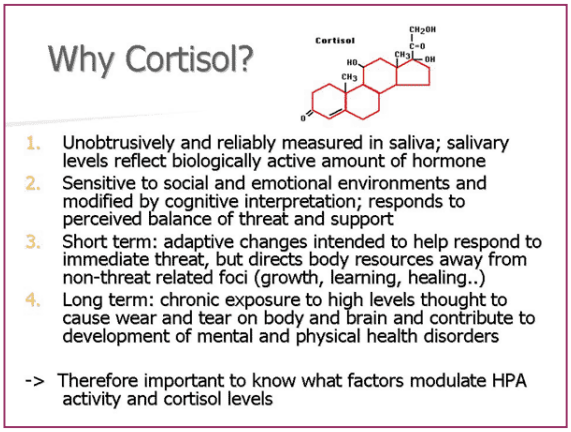


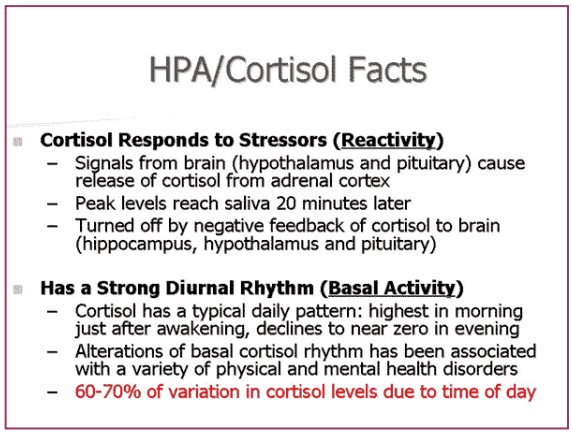
I don't think I'll spend much time going over the two major stress systems, but I do want to make the point that they are not the same -- it's not just that we're measuring cortisol and epinephrine and norepinephrine because they all in some way measure stress or stress hormones. We prefer to call them stress-sensitive hormones in that they do respond to stress, as well as responding to a lot of other things. But you can also separate these two systems based on the psychological factors that activate them and the time course of their activity. So whereas the sympathetic-adrenal-medullary system (SAM) is faster acting, it's also a shorter-lived response to a stressor, and it tends to get set off more easily in response to any of a range of events that might be challenging, whereas the HPA axis is slower acting, it comes in as a reinforcement to the SAM response, but it's longer lasting. So the time course is different, and, in addition, it tends not to be activated as frequently. It tends to come into play when the level of challenge is upped a notch to fall into the territory of being something threatening, such that the environmental challenge exceeds the resources the individual has to deal with the challenge. So there is a psychological dimension, that when threat in the environment or internal threat outweighs one's perceived ability to cope with that threat, that's when the HPA axis and cortisol is more likely to come into play. This cognitive-evaluative-psychological dimension to HPA axis activation makes it very interesting and relevant to social scientists, who are often interested in both social stress and social resources, and individual differences in people's assessments of these social-environmental variables.

Obviously there are also differences in the ease of measurement of these various systems. Epinephrine and norepinephrine are measurable primarily from blood or urine, whereas cortisol and the HPA axis is measurable in saliva. People are increasingly trying to develop salivary measures for SAM activity also – for

example alpha-amylase has been receiving attention recently as a potential correlate of sympathetic-adrenal-medullary activity. So we want to keep our eye on that and see how that does in smaller scale studies and then consider it for larger scale implementation.

So a reason for measuring cortisol, then, in addition to the theoretical reasons, it's unobtrusively and reliably measured in saliva. Another important fact about its measurement in saliva is that when it's in saliva as opposed to in blood, it reflects the biologically active or unbound portion of the hormone. So it's actually measuring the cortisol level of interest physiologically. Also, cortisol is stable in saliva for quite a long time - even at room temperature, it is stable for at least a month, probably longer. So even though we do take precautions by getting samples refrigerated or frozen as soon as possible, salivary cortisol samples are remarkably stable at room temperature.



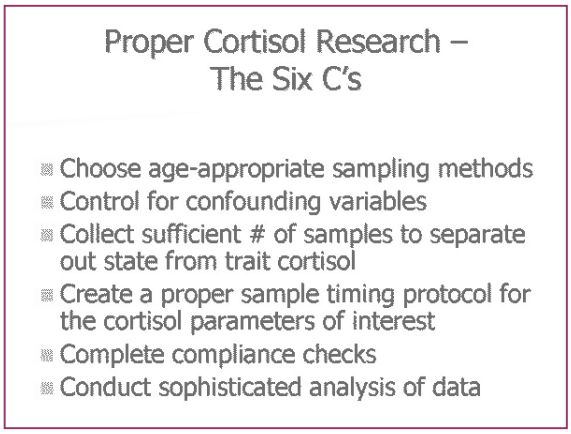


Psychologically, cortisol is sensitive to social and emotional environments, as I said, as well as to cognitive interpretation of those environments and to the perceived balance of threat and support. In the short term, it's important to know that these stress hormones are not bad. They are intended to help the organism, help the individual adapt and respond to immediate threat, but in doing so, cortisol does direct physical resources away from non-threat-related foci. In the long-term, as Teresa Seeman and many of you know better than I, chronic exposure to high levels of cortisol and other allostatic load markers are associated with the development of physical health problems. Cortisol in particular has been implicated also in the development of emotional health problems, although much of the causal information that we have on cortisol and long-term health outcomes is based on animal work and there is really a lack of human longitudinal data on the influence of cortisol on later health outcomes. Given the importance of cortisol, it's important to know what are the types of factors that modulate the activity of the HPA axis and influence cortisol levels, particularly in day-to-day settings. We all know that if you take someone into a lab and scare the @#$@# out of them, they will have a cortisol reaction, but it is also important to know "what are the things that are actually activating this hormone in day-to-day life?", and thus potentially contributing to the development of allostatic load and long-term health problems.

A few more facts. We all know that cortisol does respond to stressors, that is, there is a reactivity component. The peak levels of cortisol reach saliva approximately 20 minutes after the stressor; however, there is individual variation in that. But importantly, and the thing that makes it complicated for all of us to measure cortisol, is that there is a strong diurnal rhythm to this hormone, the basal activity of this system varies with time of day.

Cortisol is high in the morning, highest about 30 to 40 minutes after awakening. There's an increase post-awakening called the cortisol response to awakening, and from that point cortisol declines rapidly in the first few hours, and then more slowly, to near zero levels in the evening. This diurnal rhythm isn't put there just to annoy us and make it hard for us to measure cortisol. It does have a function, and it is potentially a variable of interest. Alterations in the shape and the strength of this basal rhythm have been associated with a

variety of physical and mental health disorders. More specifically, a flattening of this diurnal rhythm is thought to be an indicator of chronic strain, although, again, that has been found cross-sectionally and people haven't actually followed the emergence of this flattening over time with exposure to strain -- that is something that I'm going to try to do in my research. Very importantly, when we're talking about parsing variance, 60 to 70 percent of the variation in cortisol levels is due to time of day. If you measure cortisol across the whole day, 60 to 70 percent of your variation will be accounted for simply by time of day.



I came up with this last night in the wee hours of the morning, so if it's a little silly, I apologize. The six C's of cortisol research: Choose age-appropriate sampling strategies; control for confounding variables; collect sufficient numbers of samples to separate out state from trait cortisol variation; create a proper protocol for data collection so that you can capture the parameters of interest to you; complete compliance checks; and conduct sophisticated analysis of data. Don't worry. I'm going to go over each of these in turn. You weren't supposed to just get it from that alone.

In terms of age-appropriate sampling methods, you do have to vary your sampling techniques based on the age of your participant. I know many of you are interested in aging, and I haven't actually thought about whether there are modifications that need to be made in older adulthood, although it is potentially of concern. In young children the concern is choke hazard. You don't want to be handing a two-year-old a small cotton roll that is a perfect size to lodge in their esophagus. You want to use a long string of cotton that at early ages the parent can use as a swab and at later ages the parent or kid can hold firmly to the end of it while it is in their mouth absorbing the saliva. Stimulant use to increase the quantity of saliva in the mouth is very controversial, yet those of us who work with kids argue that it's sometimes necessary to get sufficient quantities of saliva. So if your choice is getting no sample at all, or getting a sample that may have a small amount of bias in it based on the use of something like a few Kool-Aid crystals to stimulate salivary flow, I would rather have the slightly ever-so-biased sample. Unfortunately, many reviewers remember that somewhere they read that stimulus use is bad, and so wherever they see it in an article, they just say, "I heard that stimulant use is bad, therefore all of your data are invalid", which is simply not true. There are several recently published studies showing that use of controlled quantity of certain stimulants does not significantly impact the validity of the data. For older children, adolescents, and adults, to avoid really getting anything between your sample and the assay, the easiest method is to have people just spit into a tube. I really like that method - you see my really sophisticated supplies depicted on the slide there {picture of a tiny straw and a small scientific vial}. I bought the straws at Office Depot, and the vials are sterile cryogenic polypropelene vials. I like this method because it's more portable than some other methods, so you can fit more samples in a pouch. Salivettes, for example, can get quite large and bulky. Also, in our work with the community for the Community Child Health Network Study, we've actually taken these various materials out to community members and got their reactions, and they really found the Salivette to be too medical. They reacted more negatively to that than just the plain old straw and vial. So that's something else to consider. It might be worth checking how your potential participants respond to these various methods before you settle on the exact sampling method.

Appropriate controls: medical and lifestyle variables. You cannot just collect a biological variable without knowing some of the medical and lifestyle confounds that might influence levels of that variable. I didn't give an exhaustive list here, but there are two different things you do with such information. There are reasons why you might eliminate participants from the study or from analysis and then there are reasons that you might statistically control for effect. So the case of serious illness or acute illness episodes, presence of endocrine disorders, being in the third trimester of pregnancy, or taking steroid-based medications, those are pretty much grounds for elimination in a study that is looking at cortisol levels. Other influences, such as menstrual timing, body mass, exercise levels, caffeine and nicotine intake, and medication use are things that you can measure and statistically control. Some of these are chronic ongoing things, and there are some things you want to measure immediately in the proximity of the sample: Did you have a cigarette in the hour prior to taking this sample? Did you have vigorous physical exercise in the hour before this? Did you eat a large lunch in the hour before this? There's a postprandial surge in cortisol levels, so you would want to know timing of meal times. So all of those are complicating factors, but if they are measured and statistically controlled, then they are not contributing to the error in your measure.

LINDAU: About that last point, is that self-reported, the menstrual timing, BMI, exercise, or has it also been correlated with the objective measures for those things?

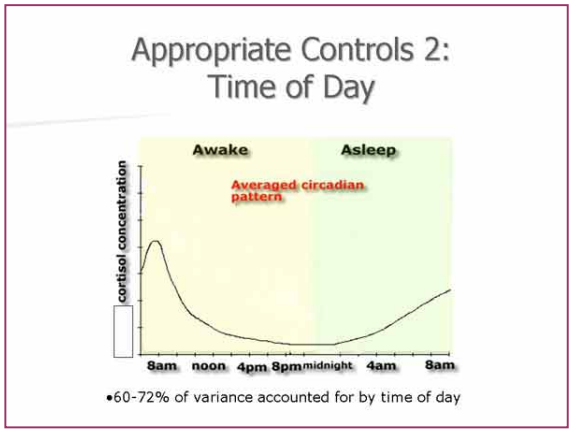
ADAM: You could, and that would be even better. I have actigraphy on my participants, so I can look at the time course of their physical activity in relation to the time course of their cortisol levels, but I'm relying on self-report for most other things.

LINDAU: It would be interesting to see if the relationship is different between the self-report and the objective.

ADAM: Absolutely. We can measure cotinine in the samples, too, and we may do that.

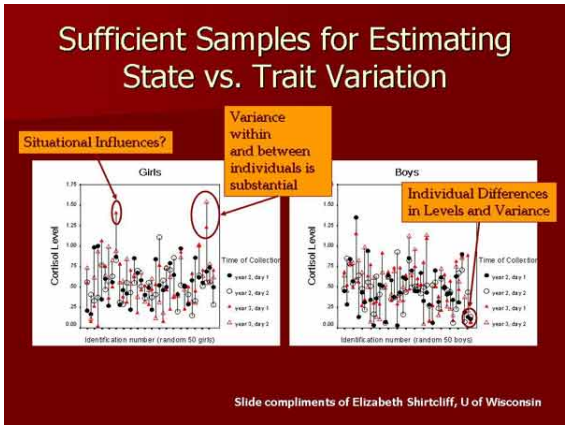
LINDAU: The BMI, exercise, medication use, nicotine?

ADAM: Yes, absolutely caffeine and nicotine intake, yes. There is an ethical issue I think we should all talk about sometime during this conference, not during my time because I've got a lot to say, but this issue of what you have gotten consent for. A lot of people are saying, "Oh, well, then later maybe I'll measure . . ." -- this, this, then that and this and that -- "in my samples." Well, if you haven't got a priori consent from your subjects to measure those things, you're ethically obligated to go back and reconsent them. So there's a fine balance between consenting them widely enough



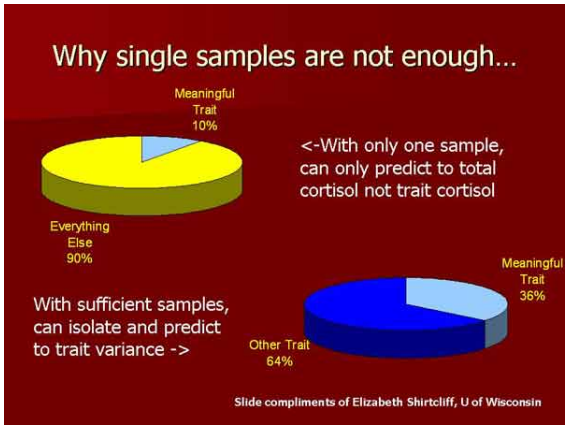
that you can measure a few additional things versus actually violating the agreement that they have made with you in measuring. So I actually consented them to measure stress-related hormones, so that I can measure DHEA as well as cortisol and maybe even testosterone without having violated my consent agreement.

Okay. So back to time of day, as I said, I've done probably four or five different studies and different samples, and the variation accounted for by time of day is always somewhere in the 60 to 75 percent range. This is a depiction of the idealized diurnal rhythm. Most actual diurnal rhythms don't really conform to that perfectly.



The point that I want to make is that from any one sample of cortisol, there are multiple influences on that sample, the first being time of day, but then there is also a situational component or "state" component to that sample, and hopefully there is also a "trait" component to that sample -- there is variability in cortisol levels within the individual and there is variability in cortisol levels between individuals. If possible, you want to get rid of the within-individual variability when looking at between-person or trait differences, or at least model state variation it in order to separate it out from whatever stable aspect of that hormone you're trying to estimate. There are a couple of different approaches that have been used to do that. Here I want to give great credit to my colleague, Birdie Shirtcliff, at the University of Wisconsin, who has done really great work applying latent state-trait models to modeling cortisol activity.

So why a single cortisol sample isn't enough and also why cortisol has gotten a "black eye" and why people have had so much trouble predicting to it is that in any one sample there is some state variation, there is trait variation, and there is error variance.//edited to here. Several studies have suggested that the breakdown is approximately this. We all know that when we do correlations in social science research, it always comes out around .32. So say we get our .32 correlation, the R-squared, we're only going to be accounting for 10 percent of the variation, of the meaningful variation, and –



HAUSER: When you refer to state, are you talking about the diurnal?

ADAM: Or you can actually just take cortisol samples in the morning, across a bunch of different mornings, and this is what Birdie has actually done in her data. So I've taken six different morning samples, and in those six different morning samples, 70 percent of it,

approximately, will be varying from day to day and about 25 percent of it will be stable from day to day, and so that's after controlling for the effects of time of day, sure.

SASTRY: Can you say something about threats and the effects that they have? Do you collect information about, you know, an actual response to a trigger, a cortisol spike in somebody during the day, and how big are those spikes compared to, say, the morning spike?

ADAM: I'll show you that in a minute because I do have data that does that, but for now I'm showing you that without information on what's going on, there's methods to find -- say you're interested in cortisol level, morning cortisol level sort of as a trait component. You're wanting to know is this person a person with high morning cortisol or low -- basal, we're modeling basal activity right now. The next method, we'll model reactivity. With only one sample, you're going to be trying to predict that little piece of the pie right there, the 10 percent. That's the amount of meaningful variation for your trait of interest. Whereas if you are only predicting to the trait variation, if you're predicting to the trait variation, you'll have a much better chance. You'll be trying to predict to a larger piece of the pie. So how do you do that and –

KURINA: Can you tell me what you mean by the meaningful trait versus other trait? Back up one more slide.

ADAM: So you've got cortisol level, like morning cortisol level, and, say, neuroticism, and you don't expect it to be perfectly associated, so the portion of morning cortisol that is attributable to whatever you're interested in is going to be smaller than the whole amount of that trait cortisol variation in the morning, and that is in turn a smaller piece of the pie. If you add in all the error variance to that picture, you can really see that the meaningful portion for your variable of interest is very small if you predict to a single cortisol sample as opposed to extracting the traits as variance.

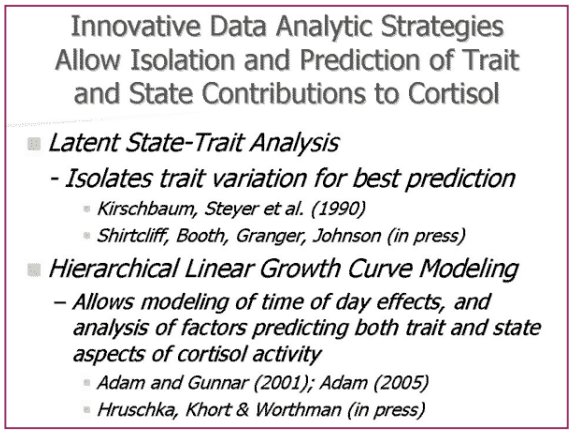
KURINA: So any kind of effect of chronic stress, which I feel has not been demonstrated in the literature, that's in the state variance, if anywhere. I mean neuroticism, you're sort of arguing for something that's more –

ADAM: It adapts over time, so change in trait might -- it gets complicated. I mean think trait. Think of the sort of stable-stable. So the chronic stress, you're right, it would fall in the state component.

CACIOPPO: This 70-percent state variance, you described it as variability across days in the morning?

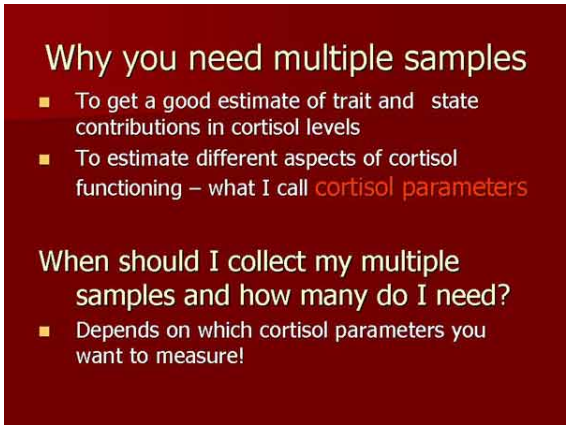
ADAM: Yes, you can do that, and you can model it.

CACIOPPO: Well, is that stable across individuals? Are there individuals who show a lot of lability over time? So is there some trait variance in that state variance that you described?



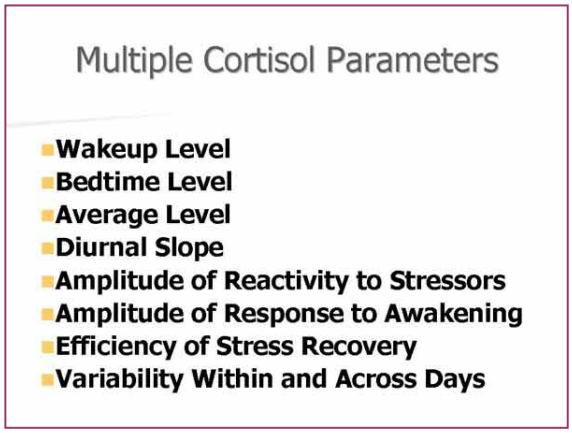
ADAM: Yes. We're going to talk later about what are actually good indicators of cortisol dysregulation. It turns out high cortisol, low cortisol, not really a sign of dysregulation, but on variability, high levels of variability from day to day is a better indicator of some sort of physiological dysregulation of the system than some of those other parameters.

Okay. So there's two methods that help you isolate this meaningful variation in cortisol: latent state-trait analysis and hierarchical linear growth curve modeling. Latent state-trait analysis I'm not as familiar with, but I wanted to introduce it to you and give you references where you can read further about it. Hierarchical linear growth curve modeling I've done a lot with. It's an approach that allows you to model the effects of time of day on cortisol and to kind of come up with a latent estimate of the diurnal pattern for each individual and then simultaneously predict individual differences in those diurnal patterns, and when I get to the actual example, that will become clearer, but you can also refer to these references for these approaches.



Just a taste of what a latent state-trait model does, it essentially uses structural equation modeling. It's similar to a measurement model. You have multiple measures of cortisol across. It could be across a couple of days in the same year, it can be across a variety of different time frames, but the notion is that you're getting multiple measures of the same biomarker and you are extracting what is the common variance for each individual and you're calling that your trait variation, and then that is what you're predicting to. Now, if you're interested in modeling the state variation, you have to do something else, but if you're interested in isolating the trait variation and predicting that, then this is -- and I would argue that this is what you should plug into your allostatic load measure in terms of your cortisol if you have sufficient numbers of samples to do so.

We'll talk about implications for your choice of measurement in a bit if we have time, but it implies a different thing about, you know, how many samples you might need and when you may choose to measure them. I'm not going to go over all these details. You can do a similar approach, actually, if you wanted to get latent estimates of the level of cortisol and cortisol diurnal rhythms. That's sort of another variant of the same approach. Across three different studies, latent state-trait studies that were done by Birdie Shirtcliff, she found relatively similar levels of trait variation in her cortisol samples. So why you need multiple samples is to get enough data to be able to separate out these state and trait contributions, but you also need multiple samples in order to estimate different aspects of cortisol functioning or HPA axis functioning, something I'm calling cortisol parameters.



So let's talk about some of the things that you can look at in cortisol. You can look at your wakeup levels, you can look at your bedtime levels, average levels across the day, slope of the diurnal cortisol curve. You can look at the amplitude of an individual's reactivity to stressors. You can look at the amplitude of their response to awakening, that 30 to 40 minutes of post-awakening sample. You can look at the efficiency with which they recover from a stress reaction, and you can look at their variability in each of these things across days. So you can see how easily I can spend years holed up in my office analyzing all of these different aspects.

What you find is that these are related to each other, actually to differing degrees, but there's also some independence. These represent different indicators of cortisol functioning, and you might choose to measure different ones based on your theoretical question. So, for example, the cortisol response to awakening has increasingly emerged as the best indicator of current strain. So average level is not as strongly associated with sort of current strain, particularly work strain. A high cortisol response to awakening seems to be associated with that. In terms of effects of chronic strain, the flattening of the diurnal rhythm is thought to be the best candidate as the marker of interest for exposure to chronic strain. You know, I could talk about each of these in turn, but it would take another two hours, but just have some awareness that they each mean different things potentially, and we're still learning what each of these means. It's going to take being precise about measuring these in studies to actually increase our understanding of the meaning of each of these parameters, but they each imply a different thing about when you would gather your samples. There is some sign that morning levels may contain more trait variation and evening levels may contain a little more state variation. So if you're only interested in wakeup levels, you might choose to measure one sample in the morning across three days so you can get that latent state-trait estimate instead of doing morning, evening, but then you lose the ability to look at the diurnal rhythm, and that's really sad. I mean there are a lot of choices to be made, and I'm sure you've all had to make them in designing your studies.

SASTRY: Does it matter how people wake up?

ADAM: Cortisol response to awakening, it doesn't appear to. There is debate as to whether time of awakening matters. There are some studies that show that it does and some studies that show it doesn't. But they have done alarm, and you can wake people up in the middle of the night, for example. It's not programmed in, you know, like the diurnal rhythm. You can wake them up in the middle of the night, they'll have a response, and you can wake them up again and they'll have another response. I don't know who volunteered for that study. (Laughter.)

LINDAU: To what degree are the claims that you're stating about these –

ADAM: Interpretations?

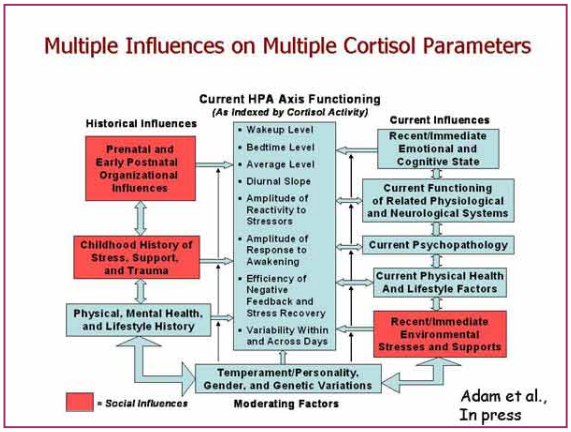
LINDAU: Yes, or how much are they understood with regard to older populations?

ADAM: I don't know. I think the study of older populations will actually tell us a lot. For example, this notion of lifelong chronic strain being reflected in a flattening of the diurnal rhythm can be better tested in older populations. But really what we need are longitudinal data so that we can watch the emergence of changes in these systems. My theory is that in the short term you have high reactivity, and so exposure to stress, you know, your cortisol levels are shooting up and down all the time, and then over time

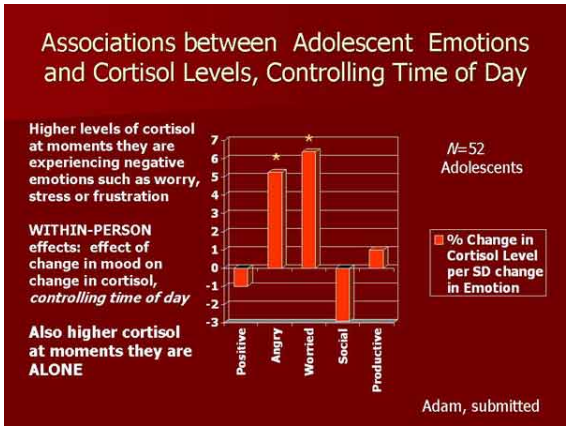
you're going to have a flattening. So in the short term you'll have higher average, but then with chronic strain, you'll have a flattening, and then eventually an exhaustion, which is reflected by lower levels. But, again, that has been derived from animal models and it has been hypothesized in humans, but nobody has followed people to actually watch these rhythms emerge or these changes emerge. So I would feel much more confident in interpretation once we have a little longitudinal data. And I'm doing that. I have a sample in which I'm measuring all of these parameters over a period of four or five years, measuring them yearly in relation to life events, et cetera, et cetera. It's a younger population, but it would be certainly interesting to do, and John's group will have those data, similar data in an older population.

SEEMAN: The cortisol data that I'm aware of for older adults mostly suggests that the biggest shift is in that evening nadir and that they don't turn the system off like they used to and that the ones who show the biggest increases are at highest risk for the negative outcomes, but there's not as much salivary data on them as there is for younger groups.

ADAM: Yes, and that would be really interesting to look at in relation to changes in sleep that occur. On my same sample, I have actigraphy and I'm looking at also sleep timing and quality. This is a population at risk for depression and anxiety disorders, so I should see some similar changes in that group.



Okay. This next point is to scare you even further if you're not scared enough already. This is to say that not only are there these multiple parameters that you can look at that may all mean slightly different things, there has in fact has been a lifetime of influence on this system -- factors starting from the prenatal period through to the present. All of these factors are impinging upon and influencing this system. And my argument is that you can just go to current psychopathology or current stress and measure average cortisol and get some sort of meaningful association -- in fact, it's a miracle if you can do that without taking into acc ount at least some of the things like current physical health and lifestyle or recent environmental events that may be activating the system in a normative way. So it's hard to call a system dysregulated when it's high or turned on in the presence of severe life events. So after an earthquake, it would be a bad thing if you didn't have higher cortisol levels. So actually starting to measure and either control for or model these other variables will hopefully allow the meaningful variation that you're particularly interested in to emerge. I'll give you one example of a study in which I've done that, and any comments on this model, you know, are welcome. It's kind of the kitchen-sink model, but there are studies to support every link that has been drawn in this model.



So this is my adolescent daily experiences and stress study. This is actually a normative sample and then I moved to a risk sample after that. I was interested in what experiences in the daily lives of adolescents influence their cortisol level. The study is actually going to give an example of the fact that you can model state variation in these levels as well as trait. So I randomly beeped adolescents seven times a day for two days. Actually there was a wakeup and a bedtime sample that were fixed, and then they were beeped within that time point, and the wakeup plus 30. When they were beeped, they completed a diary report of their location, activities, thoughts, moods, and then 20 minutes later they provided a sample of cortisol. This is an example of the diurnal rhythm of all the data smushed together, which makes it look prettier than it is, but even in this, you can see to a large extent the rhythm is present, but there is certainly some variation. And the question is, when you do see variation from the typical or expected diurnal rhythm, is that variation predictable by aspects of the immediate or momentary circumstance. So what I do to look at this is, for each individual, separately, I used HLM growth curve modeling to model each person's diurnal rhythm, and then when values are above or below the expected value for that individual for that time of day, I asked the question: Is there something about the circumstance that predicts that? In particular I was interested in the ups and downs of adolescents, so as adolescents' moods go up and down, can we see variations in the hormones that correspond to those mood changes, and indeed we could. It took a lot of time. You have to properly model those diurnal rhythms. You can do quadratics and splines and all sorts of fun things, but it's important to get that right. Again, we could do another hour on that. But once you control for the effects of time of day, you do see that increases in negative emotions, both anger and worry, were associated with increases in cortisol levels. And that's a within-person effect, so within-person changes in mood, predicting within-person changes in cortisol levels. You'll notice that there was a trend for feeling social to be associated with lower cortisol levels. Also, there was a finding that when kids were alone, when younger adolescents were alone their cortisol levels were higher, and this effect went away as they got older. I thought that was an interesting developmental effect.

This is actually now moving to another study with similar methodology but with a risk sample. Here is where I'm trying to sort out some of these different time courses of influence, and this is now looking at the elevation of their cortisol levels, elevation of their basal cortisol levels, and found that their negative emotion on the day of testing was associated with higher basal levels on those days. That shouldn't be too surprising. This is using Connie Hammond's life events interview. I looked at sources of ongoing chronic strain, but it's still relatively recent, kind of within the last year. Those who were reporting ongoing family strain had significantly higher cortisol levels. Those who were reporting ongoing romantic strain had higher cortisol levels. There's a bar missing. There's also a bar for depression. Those who had clinical depression also had significantly higher average cortisol levels. I know you're going to ask me whether these were independent effects, and, yes, they were. The interesting thing is, actually the association between clinical depression and cortisol was not present as a simple association. Only when you accounted for the normative activation of the system with negative emotion, recent life events, and ongoing strain did you see the association between a major depressive disorder and cortisol. So I guess I'm arguing that some of the associations that we're interested in are obscured by failure to control for the normal activity and expected activity of the system. So finding what's pathological activity of the system is difficult unless you actually recognize and model the fact that the system has a normative function and normative responses to events.

DAVID: Do you think that getting beeped might have been regarded as a threat by some of the people and maybe not by others and that might have confounded? I'm speaking as a physician who carries a beeper.

ADAM: It's possible they experienced it as stressful. I don't know. It should have shown up in their diaries. I'll have to think about that. They do report on recent stressors, and rarely do they put being beeped as the most stressful thing that happened in the last hour. Usually it's their boyfriend, girlfriend, or the SATs.

So just be aware that there's a literature on compliance with ambulatory cortisol protocols, and you'll have the minutes, you'll have the notes, and you can read it yourself. People do comply better with ambulatory protocols if they are aware they are being monitored. Now, the nice thing about this is that all you have to do is tell people that they are being monitored and they comply better. You don't have to necessarily actually monitor them. You can just pretend to monitor them and they will be better about compliance, or you can monitor a random subset or something and that would improve compliance. The compliance is actually pretty high in these studies. It is approximately 80 percent without the monitoring and then up to 90 percent with the monitoring.

There are several different aids that you can use. This is my favorite aid. I went to the Dollar Store for my study and bought a whole bunch of little kitchen timers. I did test them. You know, they're pretty accurate for your dollar. So for the wakeup plus 30-minute sample, I just had people turn on the egg timer, and when it goes off, they take their next sample. For the other samples they're beeped. This is the compliance monitor. You put the straws or cotton or whatever in there. It electronically records the time that they open the vial, so you can tell at least whether they opened the vial when they were supposed to. Whether they actually did their sample right away, we can't tell.

And then my other favorite compliance monitor is my actigraphs, which I'm using. There is concern that if people don't take their wakeup plus 30 samples 30 minutes after they actually wake up, it definitely introduces error. So you can use actigraphy to tell when people actually physically got up in the morning and then relate their waking sample to that. At $500 a pop, it's not something you're going to use in large-scale studies, unless you're interested also in the other things which actigraphy can give you, such as activity level data and sleep quality data. So we'll go over that.

This is my intro to Chris, so I have to say it, which is the importance of theory. It's really important to be familiar with the theory and the background literature and the biomarkers you're looking at. We all know what the advantages of theory are. I think the important point is that a lot of the theories that come to us from biology may be useful in understanding social phenomena, and one perfect example of this is the theoretical approach that Chris Kuzawa brings to thinking about biological phenomena and health.

So with that, I will give up the floor.

(Applause.)

McDADE: Let's do three questions and then we'll take a quick break.

(Multiple voices.)

ADAM: Where is the cost slide? I lost it. It just seems that it may be better to do a subsample, and then cortisol is the case where in some cases it might be better to do a subsample thoroughly, having multiple samples rather than doing one sample on everyone. There are other biomarkers where one sample is fine, but cortisol is a case where you might want to choose quality over quantity.

Questions?

KARLAMANGLA: You said that the salivary cortisol lags behind the plasma unbound cortisol by 20 minutes?

ADAM: Lags behind the psychological event by 20 minutes, so the perception of the stressor. It takes about 5 minutes from plasma to saliva.

LINDAU: I'm convinced that there is a lot to be learned from cortisol. I think you gave a great talk. It was fascinating. I'm much more enlightened than I was when I came in.

Are we always going to be dependent on saliva, or is there some less invasive way one day down the road, some easier way to get at this? Do you have any ideas about that?

ADAM: To get at cortisol levels?

LINDAU: Yes.

ADAM: I mean except insofar as it affects other systems and, you know, heart rate, et cetera, et cetera, but that's not measurement. I can't think of what would be easier than saliva. Thoughts?

HARRIS: How about a patch, Emma?

ADAM: Oh, I don't know about the patch. Is it an integrated measure?

HARRIS: It's pretty new. You would wear a patch throughout the day and then it would get the cortisol secretions.

ADAM: Right. So, again, that gives you an integrated measure. It's more like urinary in that you can only get average from it. As soon as you do anything integrated like that, you're losing the rhythm information, but, you know, if you're interested in average, then maybe that's the way to go. That's interesting.

ZUCKERBRAUN: Do you see a therapeutic way to use cortisol or use something that does the opposite of cortisol or training people how to help their stress?

ADAM: I mean there's always CortiSlim.

(Laughter.)

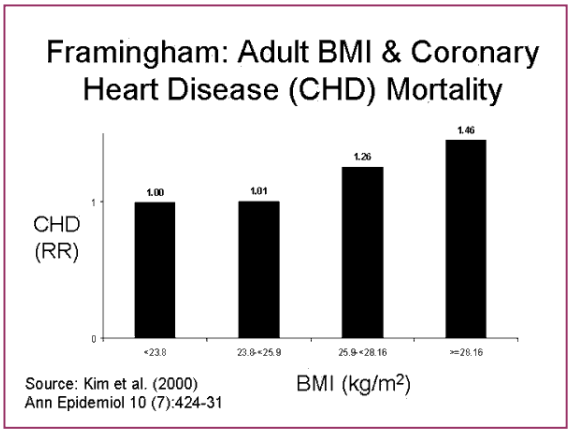
The other reason cortisol has a black eye is that some people have taken advantage of the fact that there are plenty of studies showing associations between cortisol and metabolism and central body fat, et cetera, and made a huge sales thing out of it. But in the proper hands, knowledge of cortisol levels could be used in a therapeutic manner – salivary cortisol measurement has been approved for medical/diagnostic purposes as well as simply research. I see it as a research tool, but obviously it could be a clinical tool as well. My goal in using it is to understand the long-term implications of people's social environments for their health and functioning, but I'm sure other people have other ideas.

Prenatal and Intergenerational Influences on Adult Health and Function

Chris Kuzawa

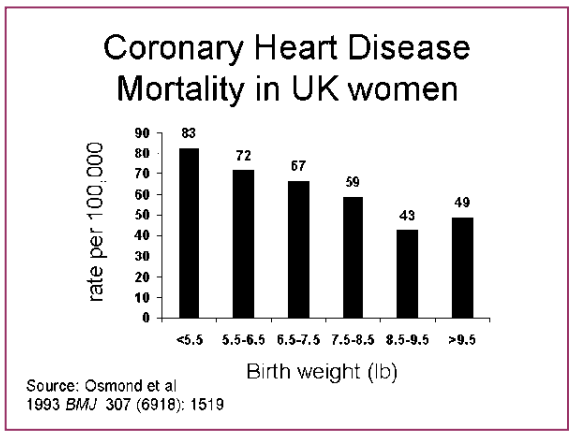
Thanks, Thom, and I'd like to thank Stacy, Thom, Jenna, and the other organizers for pulling together such a stimulating session today. I've learned a lot already. I also very much appreciate the opportunity to tell you a bit about my research today. I'm going to be talking about prenatal and intergenerational influences on adult health and function. Before I get into this, I'd like to start off with a slide that will probably not come as a surprise to any of you.

What this shows is that if you're obese, that is, if you have a high body mass index, you are more likely to suffer a heart attack. This shouldn't come as a surprise. If we know anything about diseases like hypertension, diabetes, and heart attacks, it's that adult overnutrition, eating too much of the wrong stuff, is what really increases our risk for these conditions.



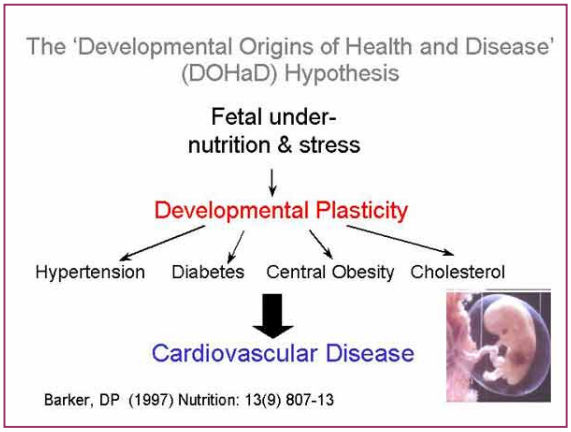
But for the same reason, this slide might come as a bit more of a surprise. This shows, again, mortality for heart attacks, but this time based not upon weight during adulthood but weight at birth, and what you see here is that the smaller individuals at birth go on to have higher risk of heart attack later in life. These data came out of Britain, from David Barker and his group at Southampton University. After finding these associations with mortality, they followed this up with other studies looking at cardiovascular risk factors like cholesterol, blood pressure, diabetes. You see the same sort of thing - lower birth weight, higher risk for all these intermediate risk factors.

So to make sense of this, they proposed a hypothesis, and the idea is that if you're a fetus and you're undernourished, as indicated in this case by being a lower birth weight baby, you're forced to adapt, and the first thing you do is adjust your growth rate. You slow your growth rate to reduce your nutritional requirements, and thus the smaller birth size, but you also end up modifying all these various systems listed here. Modifications in metabolism, physiology, and hormonal axes are an important part of this model.



When this idea was first proposed about 15 years ago, it was met with a lot of skepticism, and I think very rightly so. The original data were all retrospective, so there wasn't much information available on what happened between birth as recorded on birth records, for instance, and adulthood. The entire 60 years in between was a big black box, but since then there have several hundred

studies, I would say, in humans, including quite a few prospective cohort studies, showing similar kinds of relationships. And more importantly, there have been hundreds of studies of animal models that replicate similar findings with experimental designs.

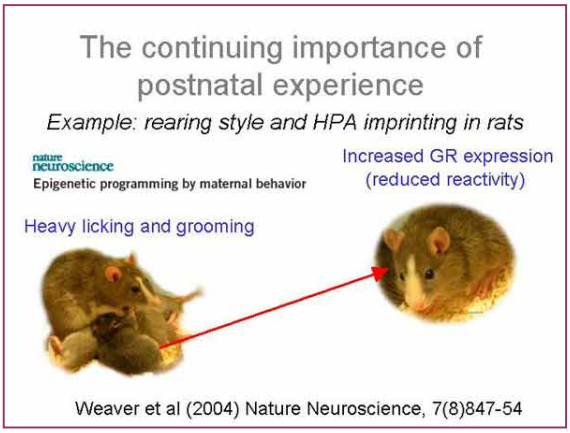


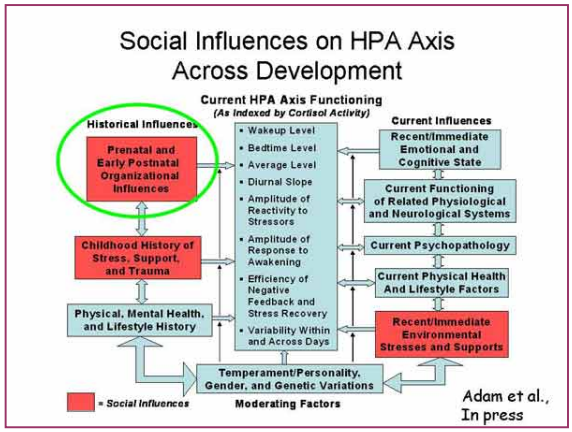
So what we're really talking about here, what this model is all about, as I've outlined here, is developmental plasticity. Developmental plasticity, in a nutshell, is the different bodies that a genome can build in response to different environments. If you take two identical twins and rear them in different environments, raise them on different types of nutrition for instance, one is going to end up taller than the other or one might actually be heavier than the other even though their genomes are identical. This is plasticity. We've known about plasticity for a long time, but what this model is doing is it's pushing that concept back before birth and highlighting the importance of these processes in utero. This research has isolated a number of different systems that are involved in developmental responses to the prenatal environment, many with long-term implications for biology and health.

Now, I want to walk through one example of a pathway that has been teased apart pretty well, I think, both in humans and animal models. In keeping with today's theme of stress hormones and the HPA axis, what this example shows is, again, the salivary cortisol response, in this case to a social stress test. This is the Trier Social Stress Test, and if you don't know what that is, that's basically a test where you have somebody go in front of an audience and have them perform a verbal task while being videotaped. So what this tells you is that my cortisol is probably up here somewhere right now. But what this also shows is that the elevation in cortisol in response to this stressor is highly contingent upon the birth weight of that individual. A lower birth weight individual, a higher cortisol response to that stressor. So you might wonder, well, what does birth weight really tell you? Birth weight is a pretty nonspecific marker. It reflects genes. It reflects all sorts of influences. With that said, we can actually go in and replicate similar responses in animal models.

This happens to be lambs. The exposure in this case was a moderate nutritional restriction, about 15 percent reduction in ad libitum feeding during pregnancy for the first 70 days of pregnancy. Now, you'll be happy to know that they did not have these lambs do a speech in front of an audience, but what they actually are doing here is injecting CRF, Corticotropin Releasing Factor, and measuring the hormonal response. Again, just like in the human data that I just showed, those individuals who experience fetal growth restriction, or fetal nutritional restriction, rather, have an accentuated cortisol response to this challenge.

Now, there are several things that I think are notable about this. Number one, this is not a particularly severe protocol. A lot of animal model research include very extreme protocols where you reduce nutritional requirements to, say, 50 percent or a level that is very unlikely under normal circumstances. The 15-percent reduction in this protocol, that's a pretty serious reduction, but it's more along the lines of what you might see in some human populations. The second thing is that this sample size of controls here has an N of 3 and the sample size of the fetal diet restriction group has an N of 7. So we're picking up very noticeable differences with very small sample sizes. When the exposure is isolated in a very controlled way, the effects appear to be quite substantial.





I've been focusing on the prenatal environment, but I want to emphasize, and I'm just tipping my hat to this because I'm not going to spend a lot of time with this, but postnatal experiences are clearly also very important. In fact, some of our best models of these kinds of 'programming' effects, as they're called have to do with postnatal exposures. These are data that I'll bet a lot of people in this room are already familiar with. This is out of Michael Meaney's lab up in McGill showing that maternal grooming style can have long-term effects on offspring, and the effect is to modulate expression of the glucocorticoid receptor in the brain. The glucocorticoid receptor is an integral part of the feedback system that regulates how much cortisol your body generates in response to a stressor. So a nurturant grooming style associated with more licking and grooming, and what they call arch-back nursing, upregulates expression of this receptor in offspring and thereby reduces reactivity in these individuals. Now, if you haven't seen this study, I would recommend that you go see it because it's just fascinating. It's really one of the best mechanistic studies that I've seen of this sort. They've got it down to the molecule, and they can turn it on and off. It's incredibly elegant.

To plug this into Emma's model which we just saw, she has current influences, constitutional influences like genetics and temperament, and then historical developmental influences. The processes that we're talking about here, this kind of programming by prenatal and early postnatal influences, would go in this box right here. I have to say, after staring at this model for a while, I have arrow envy, I really do. Look at this. Look at how pathetic my model is in comparison. It's incredible.

(Laughter.)

Okay. So the HPA is a pretty good example of one of the pathways that we worked out quite well to link early environments with later health and function.

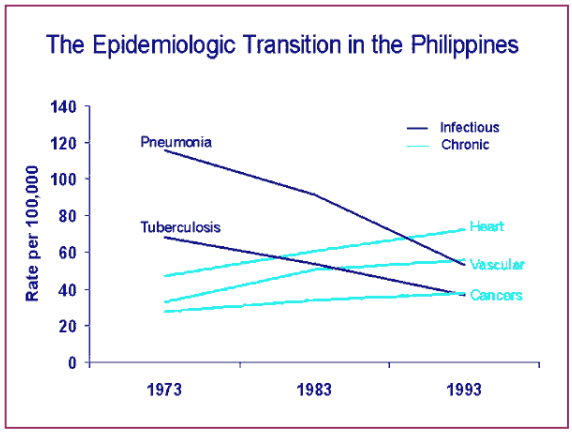
JORDAN: I have a question. The graph that you showed with the different weights for infants and the risk of cardiovascular disease later in life, were these near-term and term infants or did you also have preterm infants in there?

KUZAWA: These were Barker's findings.

JORDAN: Yes.

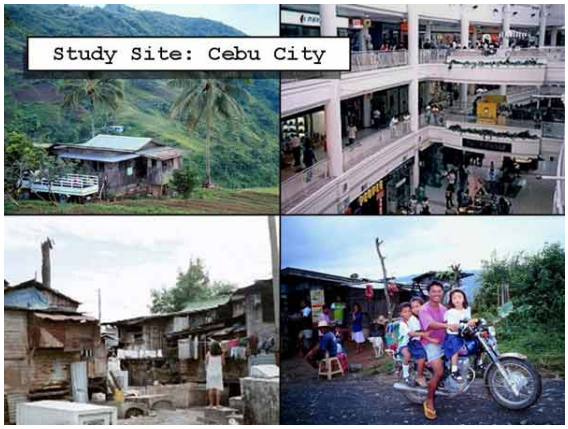
KUZAWA: I'm trying to remember now. Those were the original studies. I'm not sure if he had gestational age. I'm pretty sure he didn't have gestational age in those data. So he's just looking at birth size. The data go all the way back to the early 1900s, and I'm guessing they weren't collecting gestational age routinely at that time. But certainly more recent studies have shown similar kinds of relationships controlling for gestational age. The study that I'm going to talk about today is one example.

So back to the model. What I want to do for the rest of my talk is walk through two examples where I think this paradigm has a lot of relevance and can potentially shed a lot of light on the issue of health disparities. One is the situation in the Philippines, which typifies the conditions of a lot of nations around the world today where changes are underway.

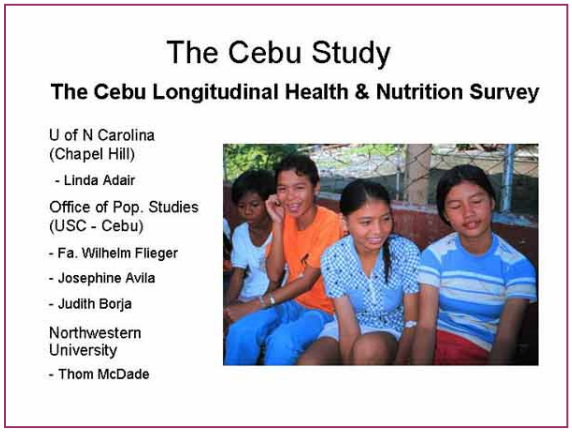


Diet, lifestyle, nutrition, disease patterns are changing rapidly, and in that situation, for instance, in the case of the population that I'm going to talk about, you've got undernutrition early in life followed by changes in diet such that individuals are then exposed to higher nutrition later in life. If this hypothesis is correct, the situation in places like the Philippines is a high-risk scenario. I'm going to walk you through a test of this idea with our data from our study and then come back at the end to the issue of U.S. health disparities, which is a very different kind of situation. Here you have chronic, stable health patterns that are very different in particular subgroups of the population, and I would suggest that this model may provide important insights into this situation as well.

So the first example is cardiovascular risk in the Philippines. Earlier Maxine told us about Taiwan, which is right here. We're now moving 100 miles south to the Philippines, a nation of about 7,000 islands. It's a big archipelago. And it's a nation that has been experiencing the epidemiologic transition of late, which is the replacement of primarily infectious disease mortality with chronic disease mortality. This is a pattern that we see over and over in different societies and it's something that is currently underway in the Philippines. Today heart disease is the leading cause of death, whereas 20 years ago, 30 years ago it was infectious diseases.

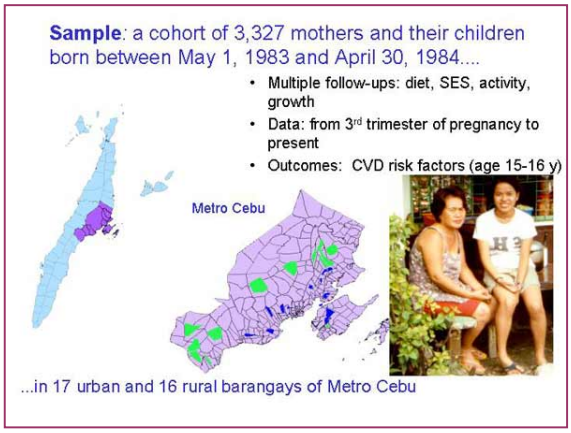


The study site is a place called Cebu, right here in the middle of the islands. It's the second largest city in the Philippines after Manila. It has about 2 million inhabitants and is rapidly growing. So it's kind of a typical mid-size city, typical of many cities in Asia. These photos give you a sense for some of the kinds of environments that our study encompasses, everything from very rural and agrarian to very densely-populated urban areas, like squatter camps, and then middle class areas as well, exemplified here by U.S.-style shopping malls filled with Kentucky Fried Chicken and people on their cell phones. So you've sort of got the full gamut here, and all of these folks are represented in our study.



I'm guessing that a number of you have probably heard of this study as it's been around for quite some time. It was started in the early '80s by Barry Popkin. It's called The Cebu Longitudinal Health and Nutrition Survey. I've listed a few of my collaborators here. Linda Adair has been the PI of this study since the late-'80s. Thom McDade is also a collaborator. We've got a number of collaborators in the Philippines who have been working on this study as well.

The design. It was started back in 1983 and it's a one-year birth cohort study. All of the women who became pregnant during that one year within 33 randomly selected neighborhoods, or what are called barangays, were invited to enroll in the study. At that time, during the third trimester of pregnancy, on average, information was collected on nutritional status, dietary intake, and many economic and household characteristics. The first follow-up was at birth when both the child and the mother were followed up, and this was followed by bi-monthly follow-ups during the first two years of life. So very detailed information on diet, nutrition, and growth were collected during infancy and early childhood. Then there were follow-ups later in childhood and adolescence, and now we're into early adulthood. The offspring are about 22 years of age, and starting to have offspring of their own. So it is slowly becoming a three-generation study.



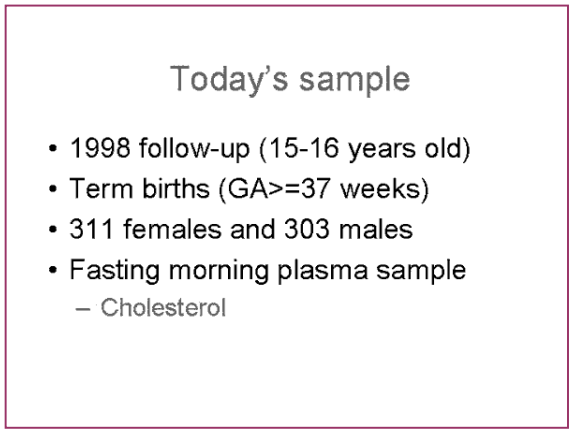
LINDAU: Those women who did participate, do you have any idea what percent they reflected of the population who

had babies in that year?

KUZAWA: You mean for the Philippines as a whole or Cebu?

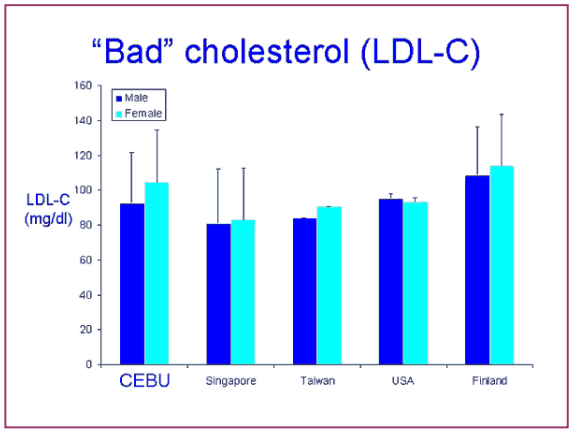
LINDAU: Let's say for that island. I mean was there a large participation rate? Was this a small fraction of people who had babies that year?

KUZAWA: Well, a number of different barangays were randomly selected, and then there was an attempt to enroll every single woman who became pregnant within those barangays. I think that the participation rate was quite high, and there were relatively few refusals. Now, the percentage of the total Cebu population represented by those 33 barangays, I'm not sure. I would have to look at the demographic records for that year and go back and reconstruct that.



Okay, what I'm going to be talking about, these are data that I helped collect back when I was living in the Philippines in 1998 and 1999. We collected blood samples and other data on about 600 males and females. This was about a third of the follow-up sample. We limited the sub-sample to term births so that small size would be the result of growth restriction rather than prematurity, as it's the former that we're really interested in. We collected fasting morning plasma samples, and in those we measured cholesterol profiles. So this is the basic outcome data that I'm going to discuss, relating it back to some of the early-life nutritional variables that I talked about and adjusting for other factors, and I'll talk more about that.

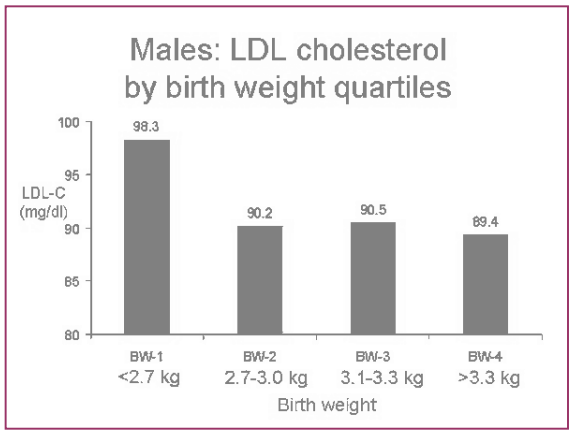
Given that we're going to be talking about cholesterol, what do Cebuanos eat? They eat a lot of rice and more rice and more rice. This thing up here is called hanging rice. They weave rice into a basket before they boil it, and it's kind of like fast food rice that you can take on the go. Big plates of rice everywhere you look. Now, not too surprisingly, for a huge archipelago of islands, fish has traditionally been the mainstay of the diet, and to this day it's still quite important, especially among lower income populations. What you also have as a traditional part of the diet is pork, and those who can afford to eat more pork, and not only that, but it tends to find its way into a lot of dishes. So there's quite a bit of fat in the diet.



Not too surprising, like just about everywhere else in the world, fast food is really taking off as well. Unlike other places, they are not flocking to McDonald's. They're flocking to Jollibee, which is a Filipino fast-food chain. Apparently in the last two years, Jollibee has been the highest performing stock on the Manila Stock Exchange. This will give you a sense for its growth right now. And what you also see more and

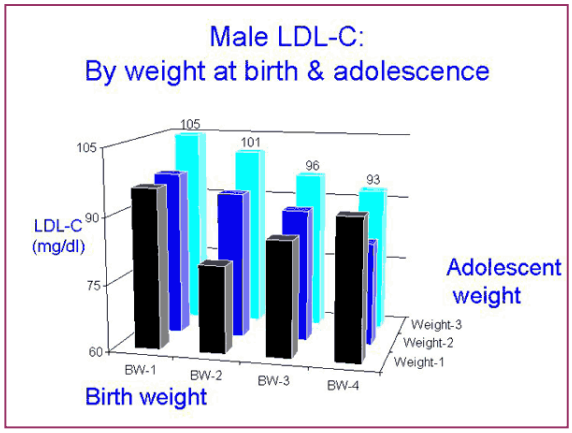
more, just about everywhere you look, are these little Burger stands that are popping up, like this one, and you can get a little burger at a place like this for about 10, 15 cents. So a high-fat diet is working its way into the culture, and that can only mean one thing: cholesterol.

So let's look at cholesterol in this sample. As I mentioned, the sample was 15- to 16-year-olds when we got these samples, and this graph compares their LDL cholesterol to adolescents in other populations. Their 'bad' cholesterol levels are actually higher than in Singapore and Taiwan, maybe even marginally higher than in the U.S. We were surprised to find this. If you look at other lipid outcomes; similarly, not very favorable. Their high-density lipoprotein cholesterol is low compared to other populations. So this is a very atherogenic lipid profile, perhaps relating back to the dietary and other kinds of factors that I talked about.



So the question that I'd like to address now: Is cholesterol in this population predicted by birth weight? This is the first stab at the test of this hypothesis, birth weight being the crudest measure of the prenatal environment. Just to give you a sense, birth weight in Cebu is about a pound less than in the U.S., on average. Again, this is a rough marker for early life nutrition.

This graph shows the raw data stratified in quartiles of birth weight in males, and, again, looking at LDL. As you can see, we do have a little bit of an elevation here at the lowest level of birth weight. It's nothing too spectacular, and it's certainly not graded, as you might want to see, but one of the questions that I think is very interesting that I kind of alluded to earlier is the idea that there might in fact be an interaction - undernutrition early in life followed by overnutrition later in life, perhaps this is the highest risk scenario.



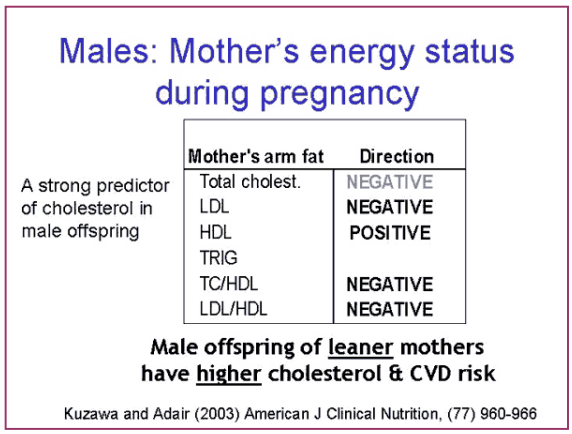
This next graph is a very rudimentary way of getting at this. Here we have the same quartiles of birth weight, but now stratifying on three levels of the child's own weight at the age of cholesterol measurement. When you look at the kids who are heavier -- I should emphasize that there's just about no obesity in this sample at this age, so we're not heavy as in American heavy. These kids are lean. But if you take the highest third of that sample, you really start to see the inverse relationship with birth weight, and if you look at the kids in the middle part of the distribution, you see the same sort of inverse relationship, suggesting that as birth weight goes down, there is in fact a rise in LDL, but that you may have to have a certain minimum level of nutritional status or weight, as an adolescent at least, for it to manifest. Because in the thinnest kids, and these kids here are quite lean, the relationship breaks down. Now, birth weight is, like I

said, a rudimentary way to get at nutrition, and we have better measures in our sample. We have maternal nutritional status and other factors.

DAVID: Did you use BMI or is that weight?

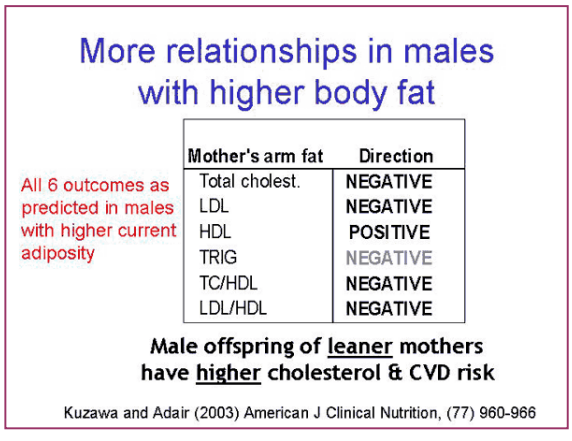
KUZAWA: That's weight, just body weight. This is stratified on weight at birth and weight in adolescence.

So the next thing I'd like to do, let's leave birth weight behind and look at nutritional status during pregnancy. We've got these nutritional data that were prospectively collected back in 1983, so we can ask these kinds of questions. What we're using here is a measure of energy stores, adipose tissue. We use triceps skinfold and arm circumference to calculate an arm fat area and use that as our marker of energy status.



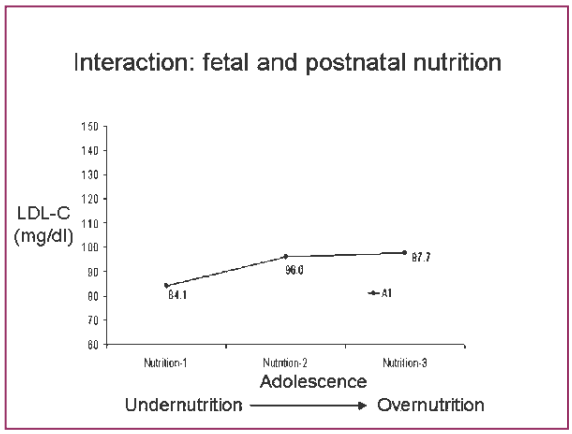
This is a little bit busy here, but what we have is LDL again, but this table shows a full panel of different lipid outcomes. Higher measures of everything on this list is bad – elevates risk for CVD, except for HDL cholesterol, which is beneficial when high. This table summarizes the outcome of six different multivariate regression models; one predicting total cholesterol, adjusting for a variety of factors; the next, LDL, and so forth. What we see is that maternal energy status during pregnancy is a pretty strong predictor of pretty much the whole lipid profile in males.

Going back to what I just showed in that prior graph, you might expect these relationships to actually become stronger if we limit the sample to individuals who had a bit more body fat themselves at the time of cholesterol measurement. To evaluate this, I ran the same models again, but this time limited to individuals in the upper half of the triceps distribution, and when you do that, the relationships strengthen even further, and now we have a borderline relationship with triglycerides. So the whole lipid panel is predicted by maternal energy status during pregnancy. We've known for a long time that the more fat you put on as an adult, the higher your cholesterol on average is going to become. But what this is showing is exactly the opposite, that if your mother has higher body fat during pregnancy, your cholesterol goes down, and we don't know that that's necessarily causal, but it's consistent with this hypothesis that Barker proposed.



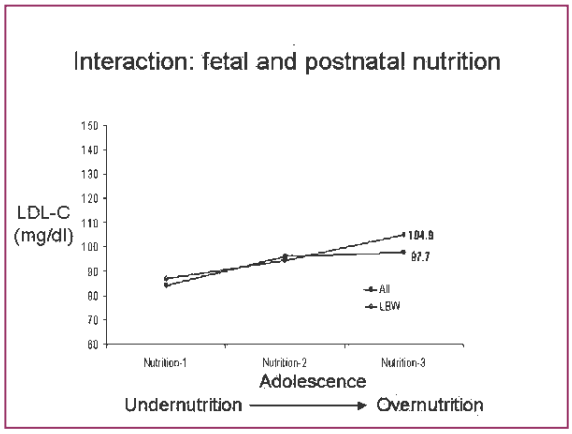
LINDAU: Do you know anything about maternal smoking history or what the patterns of smoking were like among women at that time?

KUZAWA: Yes, we do. It's very uncommon among women in this society. In fact, smoking, even among men is quite different. Most men smoke, but they don't smoke in the way that we smoke here. They might have one or two a day, because you can just buy single cigarettes on the street. It's very uncommon among women, and it was even less common back in 1983. But we do have that information and we've looked at it, yes.



Before I go on, I should mention, you may have noticed that all of those graphs were of male outcomes. Well, there's a reason for that. There's absolutely nothing going on in females for lipid profiles, but this is not that unexpected. In fact, we see sex differences of this sort even in animal models of fetal programming when the outcome is lipid profiles. So there appears to be something going on, and we don't know exactly what, with gender or with sex in the programming of at least lipid profiles. We don't see the same kinds of sex differences in blood pressure and other outcomes.

So let's move on here and see what's happening with this process in this population experiencing nutritional change. In this graph, I've got a variable on the x axis, which is a nutrition variable. It's a composite of a number of things: energy intake, dietary fat intake, adiposity, income, physical activity. It's kind of an overall measure of bad lifestyle. And, not too surprisingly, as you go across this axis, LDL goes up by about 13, 14 milligrams per deciliter.



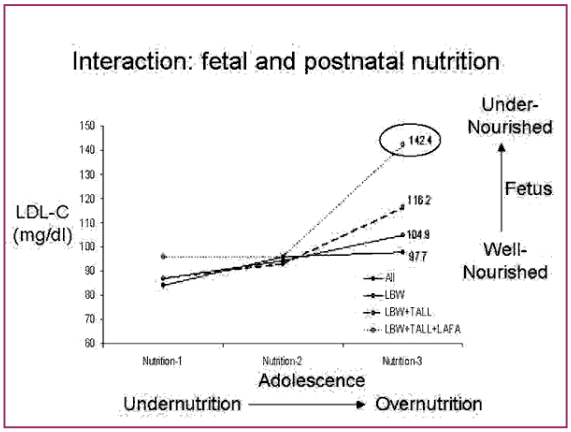
Now, I will limit this sample and look at different subsets to try to get at whether or not this increase in LDL across this variable might vary depending upon your experiences in nutrition as a fetus. First, I limit the sample to low birth weight. And this is not actually low birth weight. This is the lower half of the birth weight distribution, or a birth weight less than about 3 kilograms. If you do that, well, nothing too spectacular here. At the very highest levels of current nutrition, there's about a 7 milligram per deciliter increase, probably not significant. But, as I've been emphasizing all along, birth weight is not a particularly good measure of in utero nutrition. You have individuals who end up small at birth purely because they had a low genetic growth potential, for instance, and you wouldn't expect those people to be at higher risk.

Because we have other characteristics of the mothers, we can use these to refine our criteria a little bit. In particular, maternal stature, I think, is an important variable, because, on average, a taller woman is expected to give birth to a larger baby. We know this. And so by the same reasoning, a tall mother giving birth to a small baby is a more unusual outcome, more likely to reflect something having gone wrong in utero. And this subset is what is shown here on the graph. This is now not just the low birth weight

individuals, but it's low birth weight individuals born to taller mothers, mothers in the upper half of the maternal stature distribution. As you can see, there's more like a 20-milligram per deciliter difference here in LDL between this group and the population average. If you're on statin drugs, you would be very happy to lower your LDL by that much. That's a big, epidemiologically significant difference.

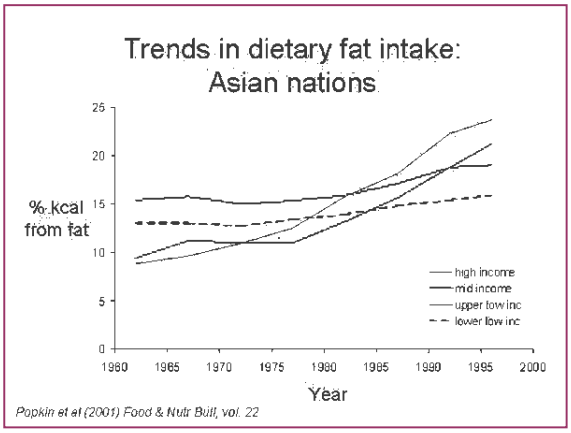
Now one final criteria. We've already limited the sample to low birth weight offspring of tall mothers. I'm going to limit it one more time to those offspring of tall mothers who were poorly nourished during pregnancy, again, as indicated by triceps skinfold. When we do that, it really takes off, and there's a large difference in LDL in this sub-group.

So several observations come out of this. Consistent with the basic premise of the fetal origins hypothesis, individuals who are undernourished early in life, as indicated by this increasingly selective criteria of fetal undernutrition, but relatively overnourished later in life, are at highest risk.



This is consistent with the expectations of this hypothesis. And this graph makes other important points. For instance, there are no differences in LDL across fetal nutritional categories at these lower levels of postnatal nutrition. This highlights the contingency of these effects on postnatal exposures. It suggests that the prenatal environment is only going to have an influence on cholesterol, at least, at higher levels of postnatal nutrition. So as the Philippines moves down a path to higher dietary fat intake, weight gain and so forth, this type of process may in fact become an increasingly important influence on who is susceptible and who is not for cardiovascular disease.

So to wrap this up, what does this mean for public health? This graph shows national level statistics on fat intake. Fat intake has been on the rise for the last 40 years across the board in Asian nations, and we know that this is going to drive all sorts of chronic disease processes. But our study suggests that data such as these—national level birth weight data—are also going to be important to consider. The experience of that lifestyle change and its effect on health may well be contingent upon the prenatal nutritional and other kinds of exposures that populations have experienced prior to these lifestyle changes.



Again, there are some caveats about our data. The relationships were really only significant in the males. We don't know why this is, if there's a real sex difference, or if it has something to do with the fact that the sample was in puberty. This is one of reasons we hope to follow this sample up in adulthood to see if the pattern is stable. But

what we're finding is that undernutrition does predict elevated CVD risk, in adolescent males at least; that the health impact of lifestyle changes, as suggested by these data, could be conditioned by fetal nutritional experiences. Finally, I think all the arrows here really point to maternal health, maternal nutrition. That's the real punchline of this work, that we have to take this seriously and worry about women during pregnancy.

LINDAU: Can you say something about breast-feeding? What were the breast-feeding practices at the time that these babies were in utero?

KUZAWA: You mean back in 1983?

LINDAU: Yes.

KUZAWA: I'm trying to think.

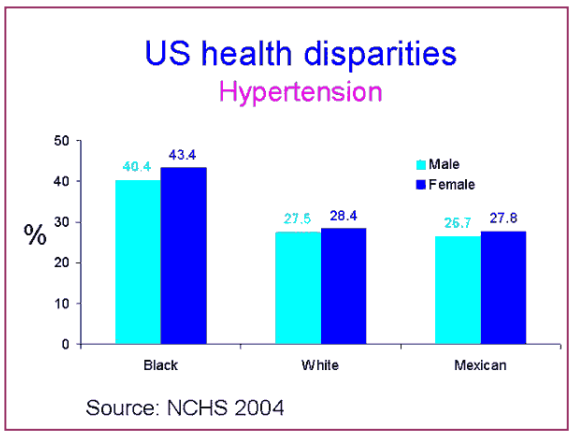
LINDAU: Postnatal, obviously, then versus now.

KUZAWA: Exclusive breast-feeding up to two months of age, about half of the sample or slightly little; by four months of age, it drops off to about, I want to say 30 percent; and by six months of age, it's rare. So there's a lot of exclusive breast-feeding at least initially, but there's a pretty quick –

LINDAU: So it seems to me that poor maternal nutrition, either less able to breast-feed or breast milk less nutritional, is part of the mechanism. I don't know if you're able to look at that in your data.

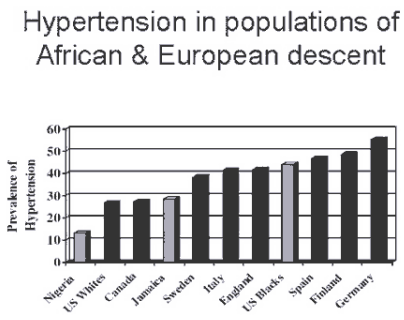
KUZAWA: In theory, we can. We have a graduate student, Elizabeth Ann Quinn, who is very interested in this, and she's begun some analyses of the infant feeding data. I think this is a really important question. And because we have this longitudinal feeding data, we can actually get at some of this. We've got the maternal nutritional status, so we can look at these kinds of interactions like you're talking about. But so far, what we've seen, and she has just done some preliminary analyses, is that there are some long-term effects of infant feeding on other outcomes, things like leptin, for instance. But in terms of the specifics and how that might play into this, I haven't actually looked at that, and I would love to do that at some point.

So moving on to example number 2, the problem of US health disparities is a very different kind of situation. This is not a situation of rapid cultural change, but instead, it's more of an entrenched, chronic pattern of difference among different subgroups.



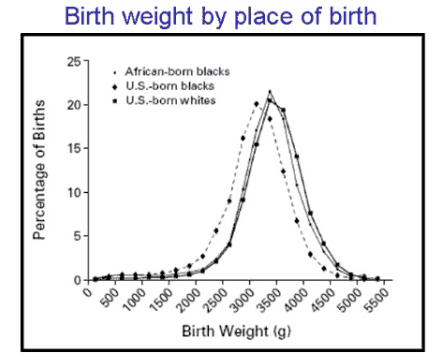
Everybody in this room is aware that there are major health disparities in this country, hypertension being one of the primary ones that contributes to higher African American mortality through a number of different pathways. In the biomedical literature there is a lot of emphasis on trying to figure out what's going on here,

and the normal approach is to perform a big multivariate analysis where you adjust for different aspects of lifestyle, and invariably what these studies find is that, even after adjusting for things like SES and lifestyle and diet, race is still a predictor of hypertension.



From: Cooper et al (2005), BMC Medicine 3:2

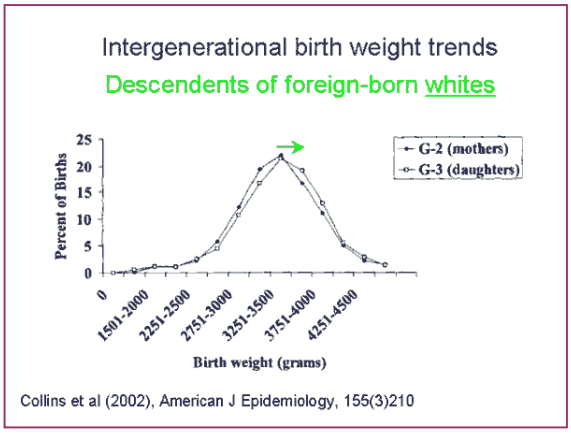
There are different conclusions that you could tentatively draw from that. The one that I would probably prefer would be that maybe we haven't measured everything in the environment to the degree that we need to in order to explain that race effect away. But when you look at the biomedical literature, there is often a tendency to kind of knee-jerk, jump at the gene as an explanation for this difference, as in this title here, "The African Gene? Searching Through History For the Roots of Black Hypertension." This implies that the roots of black hypertension are to be found in the gene as opposed to the structural racism and discrimination in U.S. society. So I think that this is a dangerous assumption to jump at uncritically. In the end we may find that there are some genetic contributions here, but I have yet to see any evidence to support this.



David & Collins (1997), New England J Med, 337:1209

What I'm going to talk about today is a different kind of model for the biology of health disparities, and one that potentially provides some insights into this problem. First of all, I think it's important to notice that populations of the African Diaspora, as indicated here by Nigeria, Jamaica, and in the U.S., have very high rates of hypertension or very low depending upon the environment, and so it really suggests that there's something very important and powerful about the environment itself that might be driving hypertension.

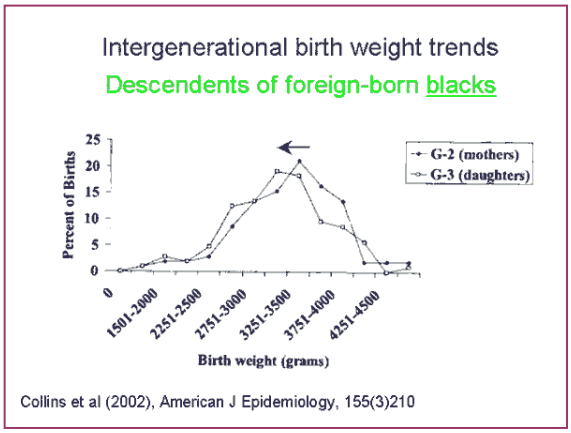
Not surprisingly, I think that there might be something going on with the prenatal environment. I think this slide in particular is very interesting. What this shows, this light dotted line here, is the birth weight distribution of U.S.-born blacks, which is shifted to the left relative to U.S.-born whites, but, importantly, it is also shifted to the left relative to African-born blacks. When African-born blacks first come to this country, their birth weight distribution is basically comparable to that of U.S.-born whites.



Collins et al (2002), American J Epidemiology, 155(3)210

These data here are from Illinois Vital Records, a big study. What they did subsequent is linked these data up with mothers' birth weight data and maternal grandmothers' birth weight data among individuals who had recently migrated to the U.S. to see if you could reconstruct the effect of the new environment on the birth weight distribution. And here is what they found. These are the descendents of foreign-born whites. The dark dotted line are the folks who have just arrived. After a generation, what you see is that distribution shifts

to the right; that is, the birth weights on average get heavier and they move towards the general US white birth weight distribution.

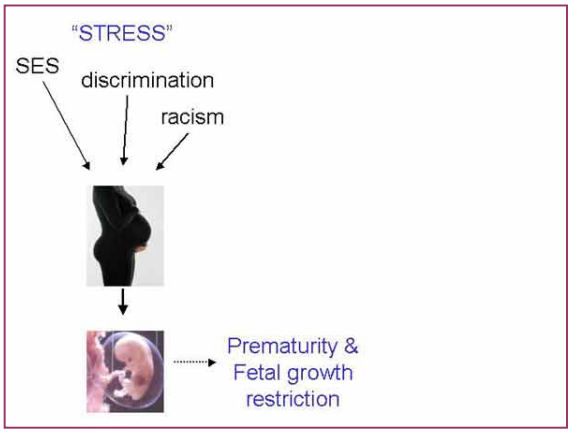


Collins et al (2002), American J Epidemiology, 155(3)210

In the blacks, however, you see the exact opposite. When they get here, the distribution is higher, and after a generation, you actually see it shifting to the left – in the direction of the US black distribution. So what does this suggest? Clearly, these are crude data here, but they suggest that there's something about the American environment that has a very different impact on the intrauterine environment and fetal growth depending upon whether you're black or white, and I think we can see that in action here.

Now, in the Philippines, a lot of the low birth weight that we had in our sample was a result of undernutrition back in 1983. There were many undernourished women in our study. I would say that's not likely the case for the shifts in the birth weight distributions that I just showed here in the US. Instead, we could probably plug in a variety of different measures of stress into our models to try to get at this. And stress has well known effects on the fetus and its development. Stress and stress hormones, notably cortisol, which we've heard about all day, actually crosses the placenta, restricts fetal growth, and is often also implicated in prematurity as well.

This prenatal exposure to cortisol can have a variety of long-term effects. In the immediate postnatal period, we know that prematurity and fetal growth restriction translate into higher risk for perinatal mortality and all kinds of poor short-term outcomes, but based upon everything that I've talked about here, we can imagine that there are going to be long-term health effects of this exposure as well.



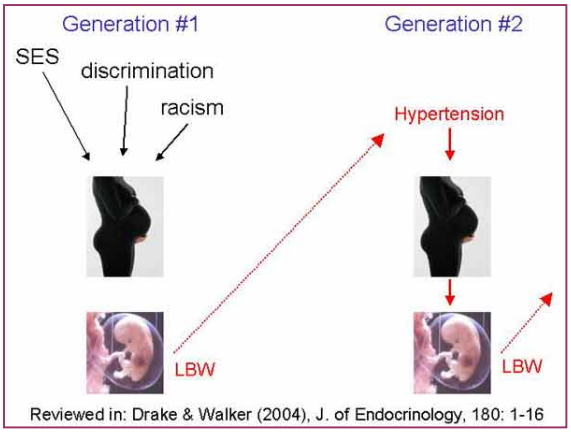
This is the Bogalusa Heart Study, a U.S.-based, multi-ethnic study. It's a long-term study of cardiovascular disease risk development in young adults. These graphs show that as birth weight goes down, we have an increase in systolic blood pressure, an increase in LDL cholesterol, and an increase in insulin resistance, which is a precursor for diabetes. So all of these long-term effects are associated with low birth weight.

KARLAMANGLA: At what ages were the systolic blood pressure and LDL measured?

KUZAWA: I'm trying to remember the design of the Bogalusa Heart Study. This study has followed a group who were originally adolescents and children, so you had a pretty wide age range, and they have since been followed into adulthood. So the age of these individuals here is probably going to be from the early 20's to about 30. That would be my guess. It's kind of like the CARDIA study but shifted to a younger age.

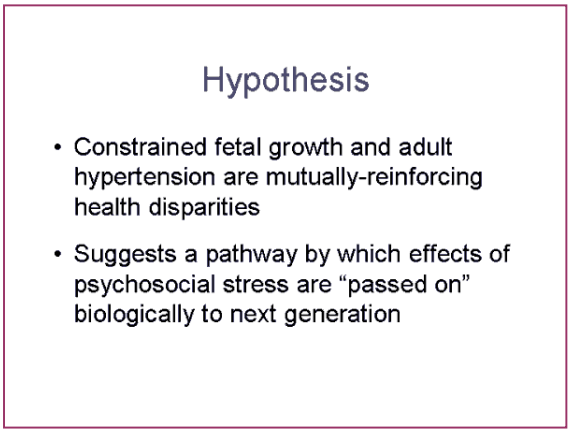
So in addition to these poor short-term outcomes, we also have some effects on things like hypertension risk in adulthood in the offspring. So you can imagine here how things like stress, discrimination, plug in whatever kinds of stressors you think are important here, could restrict fetal growth and with effects lasting into adulthood. And you may think, okay, this is sort of the end of the story, right? Some component of this adult health disparity traces back to this other health disparity; that is, the higher rates of poor birth outcomes

among African-Americans. But in some ways this is where the story begins, where it starts to become interesting, because what we know is that hypertension is a major risk factor for IUGR.



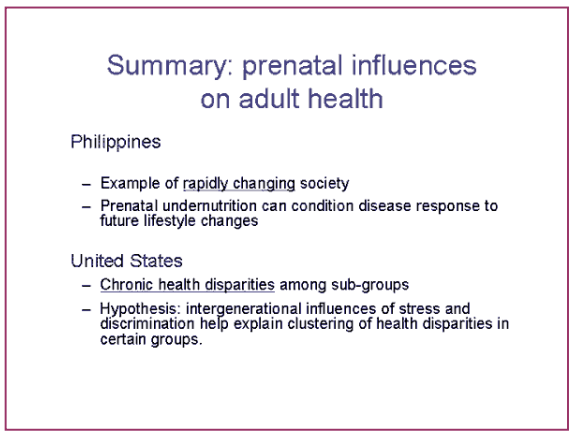
Half of the individuals who experience this adverse environment in utero are females. They go on to become mothers, many of them. So what's happening here is, we're not just influencing hypertension in adulthood, but we're influencing the quality of the intrauterine environment in the next generation in a way that predisposes those individuals to growth restriction, that is, in the grand-offspring. Of course, this is just one of many factors that's going to determine whether or not somebody becomes hypertensive or not. But hypertension increases risk for low birth weight, and you can imagine how this cycle might then potentially perpetuate itself across generations.

So this is the hypothesis. I should point out, listed down here at the bottom, nice review of this by Drake and Walker in the Journal of Endocrinology. They're not really focusing on the black/white health disparity, but they lay out this kind of intergenerational model in some detail. It's a nice review.



So the hypothesis is that constrained fetal growth and adult hypertension are mutually-reinforcing health disparities, and it suggests a pathway by which the effects of psychosocial stress might be passed on biologically to the next generation. What I think this really emphasizes is the importance of taking a life span or life course approach and also an intergenerational approach to health disparities, as illustrated in this figure. This model is life course in the sense that early life conditions are laying the seeds for later life health. Risk for future hypertension, diabetes, stress, things of this sort actually trace back in part to what was going on in utero. And it's an intergenerational model in the sense that these same adult factors feed back on the next generation through the conduit of pregnancy and the placenta and the intrauterine environment.

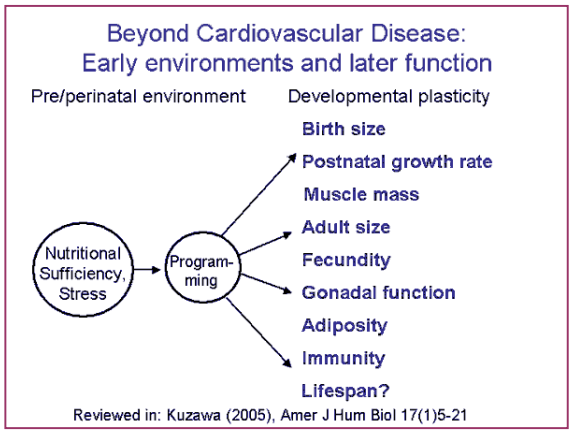
I think it's a very interesting model for a number of reasons. Number one, I should emphasize again that this is not a purely deterministic model -- it's not as simple as this figure might suggest. There are lots of other factors and influences, clearly. We don't know just how important these intergenerational pathways are as yet. But this model does suggest a way in which one might inherit a certain biological predisposition that's not genetic in origin, that traces back instead to the environment experienced by your mother and perhaps even your grandmother. We have very good evidence for this, at least for fetal growth and for several other outcomes.



So to summarize very briefly, we just walked through two different examples of where the fetal origins

hypothesis may have particular relevance. The first was the Philippines. This is a situation of rapid change, where nutrition is changing very rapidly, and the experience of that, at least among the males in our sample, appears to be very contingent on how well nourished they were in utero. And second, I think that this model also has relevance here in the US, because in some cases you have intergenerational influences that are potentially mutually reinforcing and that could stabilize into certain patterns of health disparities, including hypertension, low birth weight, and other outcomes that cluster together causally.

Before I end here, I would like to point out that there's a lot of emphasis here on cardiovascular disease, but there has been an explosion of research into different outcomes and trying to evaluate which actually relate back to the prenatal environment, and what we're seeing is that pretty much everything relates back to the prenatal environment if you take the time to look. Things like the number of muscle cells and muscle mass relates back to nutrition early in gestation. Adiposity is certainly related back to what was going on in utero. We've got very interesting data now on fecundity and gonadal function, both in males and females, good animal model data on this and some preliminary data from humans in clinical studies as well. We've got evidence for effects on immunity as well.



Before finishing here, this is a conference about biomarkers and I thought I would throw up a few ways that these kinds of hypotheses might be evaluated retrospectively. What we really need is prospective studies that are designed to test these hypotheses and that measure the exposures during pregnancy in a very specific way, but there are ways to get at some of these retrospectively; birth weight, of course, and length, which I've just talked about, but there are other ways. Many early experiences leave indelible marks in the body. Leg length relative to total standing height is a good marker of early childhood nutrition, and this has been shown to predict a variety of different adult outcomes. Fingerprint patterns, and this is a very interesting one, are laid down early in gestation, and stressful conditions changes the type of patterns that you have. Back in 1975, about 15 years before Barker proposed any of his hypotheses, there was an article in the American Journal of Physical Anthropology, kind of an obscure article, that found differences in fingerprint patterns among individuals who had experienced heart attacks. So they were actually proposing back in 1975 that perhaps there's something antenatal going on that's influencing cardiovascular disease. And, of course, we have antenatal records from clinics when this information is collected, and migration studies as well. So there are ways to get at and test these ideas retrospectively.

Our plans. As I think I mentioned, we're continuing on in the Philippines. In fact, we're in the middle of collecting data right now, and the offspring are about 22 years old. Then there is the CCHN, which a number of us at Northwestern are involved in. This is a five-university community collaborative study of the determinants of early life health disparities, both in birth outcomes and early child development. So we're hoping to incorporate some of these intergenerational questions into this study as well.

Thank you. (Applause.)

LAUDERDALE: When you see different determinants of low birth weight/prematurity completely in the Philippines or in the U.S., it makes me wonder whether there are in fact confounders in both that could be common and would tie this together. The obvious one is maternal infection, which could be strongly correlated with poor maternal nutrition in the Philippines but have completely different correlates in the U.S.

KUZAWA: Unfortunately we don't have any information on that. That is actually a key part of that last study that I just talked about, the CCHN. We're going to be looking at maternal infection during pregnancy as a contributing factor to poor birth outcomes. But it's not something that we have a way to really evaluate retrospectively with our Philippine data. What we do have is the nutritional data, and we can show that nutritional intake and things like energy status, some of the measures that we're using, actually predict birth weight. So we can get at some of that with the data that we have, but yes, it would be nice to have that data, I agree.

WEINSTEIN: I have a very practical question, which is how much cooperation do you get in measuring leg length among adults in the field?

KUZAWA: I've never done it. No, this is not a measure that I've ever used, but I've seen it used in other studies. Just to give you a sense for what it's useful for, it's one of the stronger predictors in women of the size of the child at birth, suggesting that childhood nutrition may influence birth weight of offspring. But how successful are people in measuring this in the field, why do you think it would be difficult? Basically what you have to get is standing height and sitting height. I don't think it's that difficult. The difficulty is making sure that you're measuring it in a uniform way.

CHASE-LANSDALE: It's interesting that the findings in the Philippines relate to males and then the findings in the U.S. relate to females in terms of affecting the uterine environment in the second generation. Right?

KUZAWA: Well, they didn't report the information for the U.S. by sex, so we don't know that there's a sex difference in the U.S.

CHASE-LANSDALE: You don't?

KUZAWA: No, we don't. I was just showing the birth weight distribution for everyone.

CHASE-LANSDALE: But why do you think it's only affecting the males in the Philippines?

KUZAWA: I think that it's something specific to cholesterol and lipids, because, like I said, we actually see the same sort of thing in animal models of lipid programming. You can restrict the nutritional intake of a rat, for instance, during pregnancy, and you're going to see bigger effects on cholesterol in males and oftentimes you won't see any effect on the females. So there's something about male/female biology. I don't know if it's something having to do with sex steroids, but there's clearly an interaction going on there. No one has really teased it apart yet, but I'm not that surprised that we see this in our sample.

CRIMMINS: Chris, you were showing dyslipidemia in looking at LDL. Is the same thing true for HDL? Would the same group that has high LDL have very low HDL?

KUZAWA: Yes. You mean the last graph that I showed?

CRIMMINS: Yes.

KUZAWA: I'm not sure if I've done that exact analysis, but certainly we're seeing the same sorts of effects in HDL. HDL is much lower among individuals whose mothers were poorly nourished and taller.

CRIMMINS: And have you looked at any role of infection among the children and its effect on those factors?

KUZAWA: Infection? What would be the pathway there? What are you thinking of?

CRIMMINS: From infection through inflammation to dyslipidemia.

KUZAWA: We don't have a lot of infection in this sample.

CRIMMINS: They don't have it?

KUZAWA: No, there's not much infectious disease in this sample in adolescence. This is something that we could look into, and in fact, we just measured CRP. And that should allow us to get at this question.

McDADE: Did you mean infection early in life or currently?

KUZAWA: No. During adolescence, right? What was the question?

CRIMMINS: No, I meant early in life.

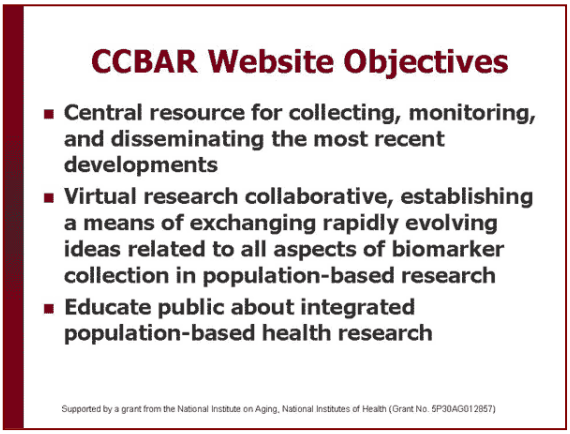
KUZAWA: Ah, early in life. We have a lot of morbidity data early in life, and we've used them to predict other outcomes. Thom has taken the lead on a lot of these analyses, looking at immune function and IgE and other factors. We're seeing an effect of those early life morbidity exposures on immune factors, but I have not looked at early infection as a predictor of outcomes like HDL.

(Applause.)

Demonstration of New CCBAR Website

Natalia Gavrilova, Ph.D. and Douglas Richardson, B.A.

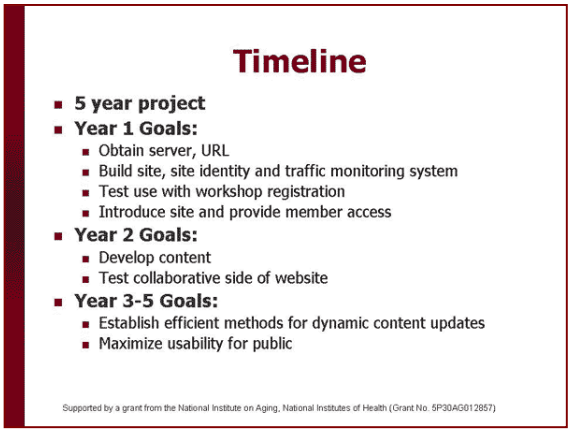
I hope you are not too tired and able to hear one more presentation on more technical issues. This is about the website. This is the website which was launched by the Chicago Core on Biomarkers, the Center on Aging at NORC and the University of Chicago.



The first goal of this website is to serve as a repository of information on the biomarkers collected in population-based settings and disseminate this information and monitor studies which are doing these population-based studies with biomarkers. The second goal is to launch collaboration on the web, which we believe is a more efficient collaboration between the researchers who are working on these topics. Finally, we would like to increase the public awareness of the studies and who is doing this research.

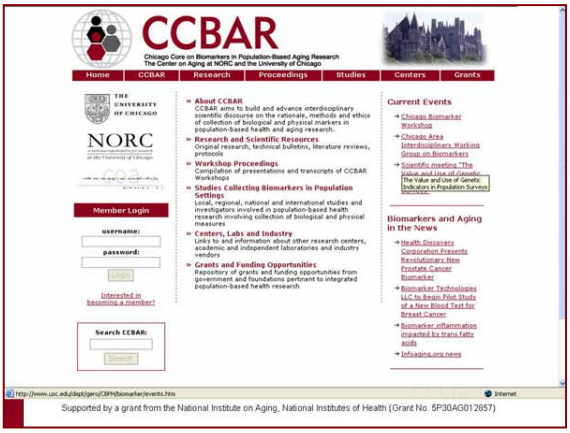
The conceptual development of this website was made under the leadership of Stacy Lindau, and we had great technical support from Doug Richardson from the University of Chicago, Department of Obstetrics and Gynecology IT Group, and the nice design of the web was done by Rebecca Stahr from Stahr Design.

This is a five-year project. For the first year the goal was just to set up the server and get it up and running and then build the site and design it and finally test the website using the online registration, which was done. And we are very grateful for your feedback about this registration site. We tried to take into account all your helpful suggestions here. And then finally we set up what is called the member site or collaborative site for online collaboration, which will be shown later. Actually we are ahead of the schedule, so for the second year, the goal of the website is just to develop content, and here we also rely on your helpful suggestions.



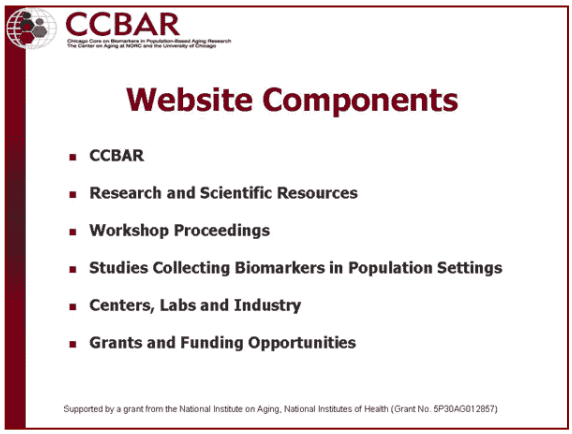
This website has two phases. One phase is the public and the second one is secure, and we call it also the collaborative site, which is accessed by the password, and the public site is publicly accessible.

The address also is very simple if you remember the word "biomarkers" and also remember that it is launched at the University of Chicago. So it is easy to remember the address of our website.



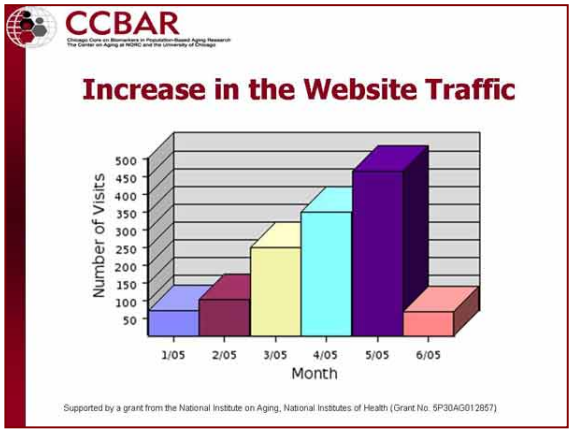
Here is what it looks like. I would like to just show you in real-time because we have internet access here. Here is how it looks. Also, it has several components. I would like to say that it's interesting that the proceedings of the last biomarkers workshop are now online, not only on the CD. If you don't like to carry the CD with you, you can simply go to the web to see the proceedings from the last biomarkers workshop.

Also, I would like to just mention briefly some sites where we rely on your helpful suggestions to build this website. First of all, research resources, which include not only the articles and books on the biomarkers, but also there are some technical reports, protocols, and if you believe that some working papers or technical reports or protocols deserve to be mentioned on the website, please let us know. Actually the website has information about the email webmaster, so you can go there, or another way is to go to the collaborative site, and there is a discussion group on the content of the CCBAR website here.



Another component is the studies collecting biomarkers in population-based settings, and here we mention some of the most well-known studies, but we believe there are more of them. If you believe that some studies are not mentioned here, please let us know. We will be very grateful to you if you inform us about more studies on biomarkers. The centers part is very small. We also believe that there are many more centers which collect biomarkers in population-based settings, so please let us know. We will be very grateful to you. Finally, the grants, and, of course, everybody knows NIH or NSF and monitors them. We also monitor these, but sometimes there are less known pilot projects, for example, which are the component of bigger, for example, program projects, and most people simply don't know about these grants. There might be smaller grants. Please let us know. We can make an announcement, like here, the announcement about the MIDUS project, pilot project. I can skip several slides now because everything is on the Internet.

Finally, I would like to say a couple of words of shameless self-promotion. I did not expect it, but our website, which was launched actually in March in its full capacity, is already found on Google or Microsoft Network. When you search for "biomarkers and aging," these words, it is on the first page already. Of course, this is only on the tenth place, but we hope that maybe soon it will be on the first or at least on the second place on these search engines.



Finally, this is the growth of traffic. Don't think that this is final. There is a drop of interest, but this is simply that June is not finished. Actually, this is all at the beginning of the month. So the traffic is not big, but we see just steady growth of the number of visitors to this

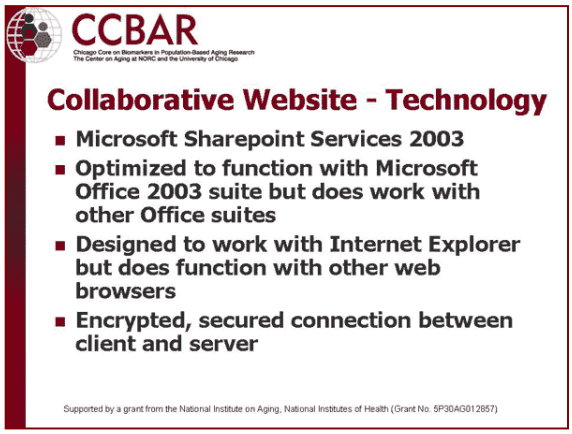
website.

Now I would like to introduce Doug Richardson who will talk about more interesting things on how to collaborate on the web.

(Applause.)

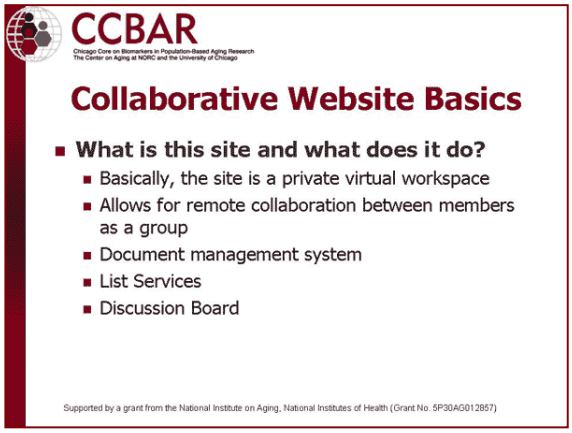
Douglas Richardson, B.A.

It's actually interesting, because the piece of the development that I was actually given was to meet the timelines that the Core developed for this project. So that being said, the collaborative component of the website was initially designed, as Natalia had described, as a place for researchers to be able to share information, create discussions, be updated on current events, design your own essentially alerts, if you will. These things should be thought of more along the lines of a big virtual workspace, so like being able to pass documents between all of us. That was really how it was described to me as to how we needed to be able to make this thing work.



So what I'm going to tell you about is how did we make it happen. We actually went through a bunch of different technological options before we chose the one that we did. I'm going to give you just some basics, and I don't know if I'd really call that a case study, but I definitely would call it an example of how you can use the website to share information, track changes, do those kinds of things.

The technology behind it is that it's essentially a Windows 2003 server. Does that mean anything to anyone? SharePoint Services is an integrated component into the operating system. So we really didn't have to buy anything else. We just had to configure it to be available on the internet. So that's something that if anyone is thinking about doing something along those lines, it's cost-effective to go with this kind of a model. It is optimized to function with the Office 2003 suite. I'd like to also let you know that it does work with previous versions of the Office suite, so you're not limited in that respect. It's just that there's some additional functionality that you get that you wouldn't get without it. It's definitely designed to work with Internet Explorer, but I haven't seen any significant issues using Firefox and some of the other browsers, Apple platform, things along those lines. It is definitely designed to have an encrypted, secured connection between you and us as you are communicating, whether that be discussions, passing documents back and forth, or whatever it is that you're actually doing on this site.

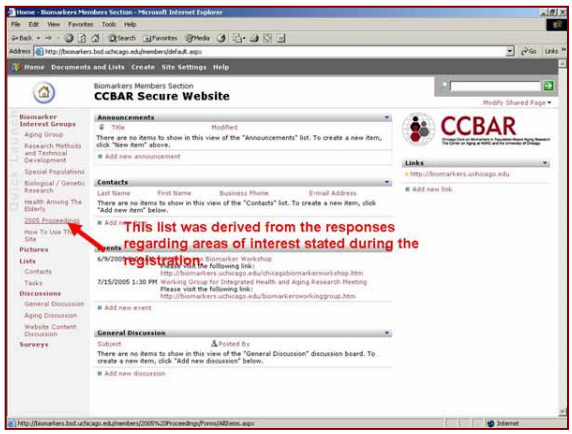


So, like I said, it's basically a private virtual workspace where it allows for remote collaboration between members as a group. No more of this emailing back and forth between individuals. If you have, let's say, five people working on a grant submission, and I know in this case what's actually research-oriented stuff, if you've

got five people working on documents, you can actually check the site, look at a version history of it, check it out, make your updates, and it will be available immediately to anyone else who wants to see those changes without having to send emails back and forth and find someone to centrally merge everything together into a final document. So it's a big help there, which is the component of document management systems. It's also designed to do list services whereby you have the ability to add yourself to be alerted to anything that happens within a particular area of interest of yours that's on the site currently. That's a big feature, actually. And you create that yourself. So it's not like someone else has to put you in that group and maintain those list services indefinitely. And also the discussion board option is readily available for people to use.

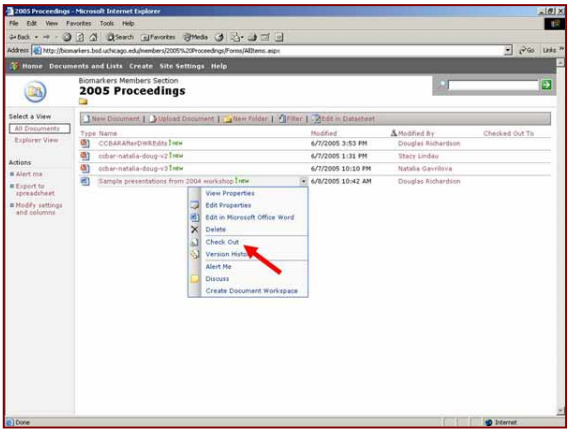
So the basics. How do we get access? Membership is currently required. This is a prerequisite that we had basically because we needed to have a controlled environment. We needed to have something that we could monitor and manage and be able to deal with the changing needs of the members without having to deal with some of the other things associated with having sites like this open to the public - permissions, the ability for people to change documents, things along those lines.

So in order to get access, everyone in this room was provided with a user name and a password. Those accounts are currently enabled and you do have the ability to use this site starting whenever you wish. To get to it, you essentially go to "biomarkers" and you click the member login and put in your password and information. This may look different if you're not using a Windows-based operating system. Put your user name and password in that we supplied and you'll see the baseline site.



So it's kind of basic right now. This is a default installation. We haven't done anything to modify Microsoft's version of this site other than to create our interest groups, which you can see on the left-hand side of the screen here. The biomarker interest groups, this is something that we actually derived this list from the registration forms. So as you filled out registrations and highlighted your areas of interest, those responses were categorized and grouped together. We initially created these. It's not limited. We can create as many as we need, and plan to. That's the intentions.

So what I was asked to do was to show you essentially how we would go through performing some of this document library functionality. In this case I'm going to choose the 2005 Proceedings components. I know I'm not actively doing this. We probably could. I'm sure the network connection is fast enough. Nonetheless, the 2005 Proceedings that you see, if we click that, you get this document library view, very limited in terms of the actual documents in it, but I wanted to kind of step you through how we go about updating a document that's there.



The first thing you do is you actually perform a check-out, and then you'll notice that when you make that click, your name will immediately show up as to who it's checked out to, so if you're working on a collaborative effort, you'll be able to notice immediately that someone is actively working on that document. It's a nice feature. The next thing you would do would be to go over and click -- and in this case it's a Word document -- that you want to edit it. This brings it to your machine. It's still stored and saved on this central file saver with version histories. So as many histories or as many edits of this document as we can possibly create, our site will actually maintain those for us, letting us know who did them at what time. It's kind of an audit trail. So if you need to go back, you can, and, in fact, you can check out previous versions of it. At that point the document will load on your local machine. We would click to track changes in this case, put in some changes, and I'm just adding a little note here to the actual live document. They're being tracked, save it, it updates it directly to the site, and then we go back and check it in. It's a pretty simple process, I guess, from my perspective, but I've been working with it. I mean we actually use this at the University of Chicago for some very large IT projects that we do divisionwide, and it has been a very wonderful tool. So after that, you check the document in and you're done and the next person can come back.

I believe you'll find out that it's time for questions.

(Applause.)

ADAM: Can you restrict who has editing privileges on particular documents or can anybody on the site come in? I mean if you want to finish my papers for me, that would be great usually.

RICHARDSON: We do currently have the ability to do that. As it is configured today, everyone in this room has the ability to check out and edit documents. We do not have that limited by owner.

ADAM: So if I put all my unfinished papers on there, then by tomorrow they'll be done basically.

RICHARDSON: Wouldn't that be wonderful.

WEINSTEIN: What happens when two people check out the same paper?

RICHARDSON: It's impossible. If you actually made an attempt to check that document out while it was already checked out to somebody else, it would allow you to make a copy, to save a copy, and it would actually create a new document in this list for you, at best. It would tell you that it is currently checked out, being worked on by so and so. "Would you like for me to alert them?" It gives you an option to do that, and it will send them an email and say whatever you want it to say. "I want my document back. Give me my document back."

SEEMAN: What happens if you forget to check it back in?

RICHARDSON: I'm assuming that whoever is working in this collaborative effort with you would be kind enough to –

SEEMAN: Somebody would call me and tell me?

RICHARDSON: Well, you can do it right through the site. Immediately, when you attempt to check it out, you can say, "I want to check it out," and alert the user.

SEEMAN: So if you didn't check it back in, it would show on here to anybody else as not having come back yet?

RICHARDSON: Yes, who has it, who is currently in possession of the editing rights.

(Multiple voices.)

LINDAU: I think it checks in automatically if you leave the site though.

RICHARDSON: No. Save it back. You can save it back in. You need to check it. It's a good practice. I don't know if you can work around through this and maybe do some backdoor action on it.

CACIOPPO: How long did it take to get this up and running? How much programming was involved, how many hours?

RICHARDSON: Almost none, almost nothing. First of all, the goals slide was a little bit understated. She said we had a year there. It was more like five months. That was to buy all the equipment, register it, and put everything together, obviously find a designer. So the first half of it is designed by Rebecca Stahr. She's a fabulous designer in the Chicagoland area. And this piece of it, we've done almost nothing to it, nothing at all. This is an out-of-the-box solution for I believe all five objectives that Stacy set for us without any programming.

CACIOPPO: You may know that NSIT shut down the web programmer's access to the server this last couple of weeks.

RICHARDSON: Yes.

CACIOPPO: How are you able to avoid being shut down?

RICHARDSON: On this site?

CACIOPPO: Yes.

RICHARDSON: Oh, we own this.

CACIOPPO: But it has a uchicago.edu.

RICHARDSON: Absolutely.

CACIOPPO: You can do it because it's –

RICHARDSON: You know, without getting into talk of the ivory tower of information technology at the University of Chicago, it's very subdivided. This device actually is not owned per se by NSIT. It's actually owned by NORC, the grant. We maintain it within the division, the

Biological Sciences Division. So that's how we get away with that. There are other sites like this in the institution, by the way, that do similar things.

LINDAU: I just want to give a couple examples of how people in this room might use it. First of all, we're going to have a much more efficient process this time of hosting the proceedings, editing the proceedings, and getting them back out, posted to the website on the public side. So those of you who have been involved in this -- like Chris Masi, we used his as an example -- in prior years know we've had to FedEx you documents, we've had to email you documents, and this will happen much more quickly. I remember very shortly after last year's meeting, Richard Suzman at NIA was calling me: "Where are the proceedings? Where are the proceedings?" He was going to a meeting and wanted to share them. This will get the information out much more quickly.

A couple other examples -- not to make examples of people, but I'm going to -- Bob Wallace and Michael David, I kind of connected them. We had a lot of collaboration for an intense period of time this year, so it would have been a great way for us to efficiently communicate, and Teresa Seeman and I have done the same, and others of you who might meet at this meeting and want to share information or start collaboration, that's really the idea. All of you could go home and start this system locally and work independently, and may very well, having seen how nicely it works, but our hope is that this will really become a way of continuing the dialogue between the annual workshops and also help us gauge interest and direction of this group as we move forward.

I want to make a comment on the contents side. Natalia really single-handedly developed content for the public side of the website over the last several months. That is really up there. Think of it more as kind of an example, placeholders, et cetera. We haven't had a lot of time to really go through systematically the criteria that would define which material ended up there and which material didn't. It's really an example that we haven't had time to focus on yet this year, but we really will during the next year. And that process of content is a dynamic process. We will have people employed by the Biomarker Core to monitor information of all kinds and continually update it, and we're hoping that this community of people will contribute to that process.

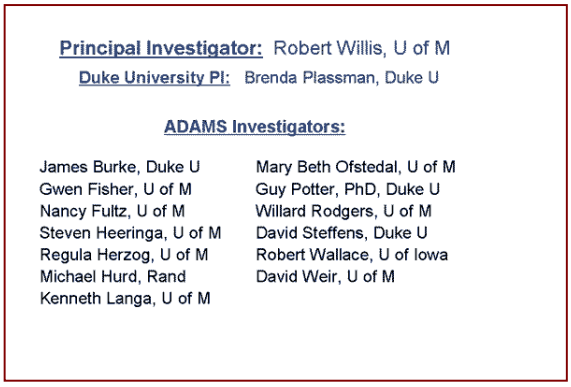
RICHARDSON: And there are some how to use documents on the web as well within the member section. There's a location where you can actually pull these documents up, and I think that one of them is -- these are actually PDF files, so as long as this has Adobe on it, you should be able to see them. You can essentially read through it. It gives you pretty much a nontechnical perspective of how to do some of the things within the members side of the site, and obviously we're open for suggestions on that.

NEW CANDIDATE FOR BIOMARKERS OF COGNITION IN POPULATION-BASED RESEARCH

The Aging, Demographics, and Memory Study

Kenneth Langa

Thanks, Stacy. So I'll try to fill Jack McArdle's shoes here. For those of you who have seen Jack speak, he always has incredibly cool and complex graphs and figures, which I don't have. So hopefully I'll say some interesting things otherwise.

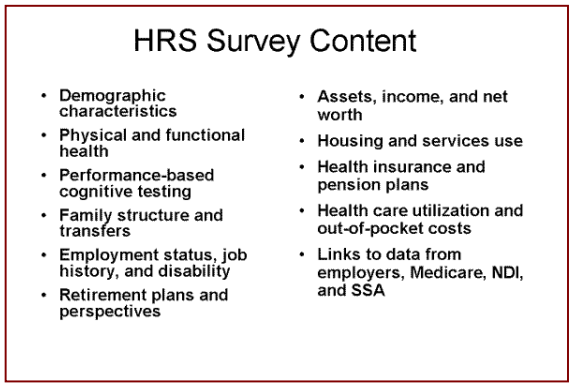


I'll give everyone an update on the Aging, Demographics and Memory Study (ADAMS) which as Stacy said I've been lucky enough to work on with a great group of investigators.

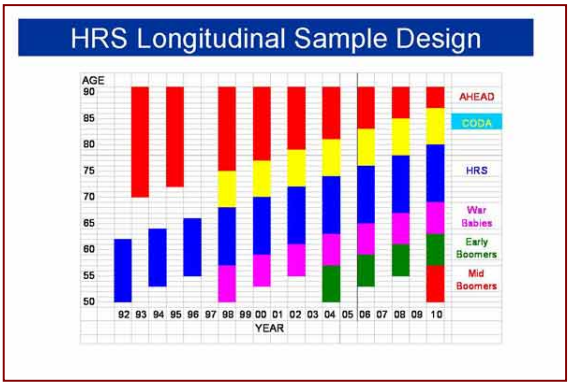
These are some of the folks who have been working on the ADAMS for the last four years now: Bob Willis is the PI; Brenda Plassman has been the Duke University principal investigator for this study. She and her team of researchers at Duke have been the ones actually going out to the homes of these people to do these assessments that I will tell you about; lots of other folks both at Michigan and elsewhere and also a great group of consultants.

What I'll do today is give a quick background on the Health and Retirement Study (HRS). I think a lot of folks here are at least relatively familiar with it. Then I'll tell you how the ADAMS was added to the HRS as a supplement, and about some of the biomarker collections we've been doing in the ADAMS, and then present some preliminary results on Apolipoprotein E genotype.

So the HRS is, as Stacy mentioned, an ongoing nationally representative, longitudinal survey, that has been collecting data every two years since 1992. Currently the sample includes about 22,000 adults who are age 50 and older. It's mainly performed at the Institute for Social Research at the University of Michigan. The HRS has been funded by NIA since 1990



Here are the general categories of data collected in the HRS: demographics, health measures, family measures, employment history, disability, pensions and retirement plans, assets, health insurance, health care utilization and out-of-pocket costs. There are also interesting links to other data sources, including Medicare data, National Death Index data, and Social Security earnings histories.



I stole this slide from David Weir. I hope he wasn't going to use it in his presentation. Here's just a quick schematic of what the HRS sample looks like. Data collection begin in 1992. The HRS actually started as two separate studies, the HRS was fielded in 1992 in adults age 50 to 61, and the Asset and Health Dynamics or AHEAD study was fielded in 1993 in adults age 70 or older. Both of these cohorts have been followed every two years through 2004.

In 1998 those two cohorts were joined and the questionnaire was "harmonized" between the two studies. Two new cohorts were selected to fill in the age range, so starting in 1998 the HRS has been nationally representative sample of all adults age 50 and older. Again, all of these study participants have been followed through 2004.

In 2004, a new cohort of individuals age 50 to 55, born 1948 to 1953, was enrolled in the study to "backfill" the aging sample. The plan is to again enroll a new cohort of 5 to 55 year olds in 2010 with a "mid-boomer" cohort.

Now turning to the ADAMS, it is the first national population-based study of dementia to include subjects from all regions of the country. Field work began in 2001 and was completed in 2005.

The objectives of the ADAMS were to select a subsample of the HRS respondents 70 and older, stratified by level of cognitive impairment on the HRS 2000 and 2002 survey. We used their performance on the cognitive measures in the core HRS survey to select the sample. We then went to their homes. Brenda Plassman and her group at Duke flew to their homes to conduct an in-home clinical assessment, a three- to four-hour assessment, including extensive neuropsychological testing.

Objectives of the ADAMS

- Select a sub-sample of HRS respondents aged 70+, stratified by level of cognitive impairment on the HRS 2000 and 2002 surveys
- Conduct in-home clinical assessments on about 850 individuals from this sample, including extensive neuropsych. testing, medical history, history of cognitive and functional change, and neurological exam
- Assign a clinical diagnosis of:
  - **Dementia**
  - **Cognitive Impairment, Not Demented (CIND)**
  - **Normal**

Objectives of the ADAMS (cont'd)

- To provide the research community with representative data to study the antecedents, and longitudinal costs and outcomes of CIND / dementia in the United States
- To use the ADAMS clinical diagnosis and accompanying HRS core survey data to assign/impute the likelihood of dementia to all HRS respondents aged 70+

One of the main outcomes of the data collection is a consensus diagnosis made by a panel of dementia experts. The panel makes a diagnosis of dementia (and dementia sub-type) based on all of the data collected by the Duke team. "Cognitive impairment, not demented," or CIND which is a hot area in dementia research. These are individuals that are not cognitively normal, they're not performing as they did when they were at their peak cognitive function, but they're also not demented. They're still able to get through the day relatively independently, but they're in this gray zone between normal cognitive function and dementia.

Other objectives of the ADAMS were: To provide the research community with representative data to study the antecedents, and longitudinal costs and outcomes of CIND in dementia in the U.S. And then also to use the ADAMS clinical diagnosis and the accompanying HRS core survey data to assign or impute the likelihood of dementia to all HRS respondents.

So we've done this very intensive data collection on a small group, small subsample of the HRS, and obtained a "gold standard" diagnosis of dementia. Using the data that's common to both the ADAMS sample and the rest of the HRS sample, we can now back impute to the full HRS sample in order to at least get a prediction of the likelihood of dementia in the rest of the 10,000 people who are 70 and older in the HRS who haven't undergone this intensive workup.

ADAMS Sample Selection

- HRS respondents age 70+ stratified by level of cognitive impairment on the HRS 2000 or 2002 core survey cognitive measures (self-report), or proxy assessment of cognitive function (proxy reports)
- Subsample of ~ 1/3 likely demented, ~ 1/3 "borderline", and ~ 1/3 likely normal based on performance in core survey in 2000 (Phase 1) or 2002 (Phase 2) for a total ADAMS sample of ~ 850 individuals

Again the sample selection was done from the 2000 and 2002 waves using both the self-respondents and proxy respondents. About 10 percent of the HRS respondents are represented by a proxy in each wave. Obviously, those represented by a proxy are the individuals who often have cognitive impairment, so you can't really do a study of dementia without using proxy reports. We chose the sample to get an expected outcome of about a third who are likely demented, a third in this borderline CIND group and a third likely normal based on again their performance in the core survey.

ADAMS Diagnostic Process (cont'd)

- In-home visit of ADAMS respondents performed by a nurse and a technician trained and supervised at Duke University
- Clinical evaluation during visit:
  - Neuropsychological test battery
  - Medical history from proxy and/or respondent
  - Chronological history of cognitive and functional change from proxy and/or respondent
  - Neuropsychiatric symptoms
  - Family history of memory impairment / dementia

The Duke assessment process is an in-home visit performed by a nurse and a technician. The evaluation includes a large neuropsych battery that I'll show you briefly; medical history from the proxy and the respondent if the respondent is able to provide information; chronological history of cognitive and functional change, neuropsychiatric symptoms, and family history, a full list of current medications, dosages and a brief neurologic exam done by the nurse.

Also we collect information on caregiver issues through a self-administered informant questionnaire that the informant fills out as the assessment of the subject is going on. For those of you who are interested in the specific neuropsych tests that are administered, this is a list of the battery here. I'll let you read that if you'd like, and then here's the rest of the neuropsych battery. The assessment process was based on work that Brenda Plassman and other colleagues at Duke have developed in other population-based studies including the Cache County study in Utah and the Twins study of dementia, also.

Turning to the biomarker collection in ADAMS, there are two "official" biomarkers in the sense of what we've been discussing over the last day and a half. We do a blood pressure measurement and a buccal swab DNA collection to obtain the Apolipoprotein E genotype. Most of you probably know that ApoE e4 genotype has been shown to increase the risk of Alzheimer's disease and dementia in many studies, so that's the reason for collecting it. We then banked the rest of that DNA sample for possible future use, but the ApoE is the only genetic information that we've actually done the analysis on so far.

ADAMS Diagnostic Process (cont'd)

- Clinical evaluation (cont'd):
  - Full list of current medications, dosages, when medication was started
  - Neurological examination
- Other data:
  - Self-administered informant questionnaire regarding caregiver time and strain

There's also something in which I'm actually quite interested. There is a seven-minute videotape of parts of the neurologic exam. You actually see the person getting up from a chair, doing various neurological tests, walking, seeing if they have Parkinsonian-like symptoms and information like that.

You also get a sense of what their home looks like because you get a quick snippet of their living room, although the environmental information included in the video is not done systematically. but I think there may be some interesting things that can to be done with this videotape information.

Biomarker Collection in the ADAMS

- Blood pressure measurement
- Buccal swab DNA collection (Apo E Genotype)
- 7-minute videotape of parts of the neurological exam
- Medical records (brain imaging and other reports)

ADAMS Field Work

- ~30 in-home assessments per month between August '01 and December '03 for final ADAMS sample of 856
- ADAMS in-home assessments performed in 42 states
- 18-month follow-ups completed for 252 individuals (29%), most of whom diagnosed with CIND at initial assessment
- Response rate: 56%, net of mortality (13% of those chosen for sample died prior to assessment)

We ask folks for permission to obtain medical records from their physicians so we do have not direct brain imaging, but we have reports of brain imaging that have been done. In addition to the field assessments on 856 individuals in 42 states, we also did 18-month follow-up assessments for 252 of the 856, mainly on those individuals with CIND. Response rate was 56 percent; net of mortality. We found early on that, not surprisingly, these folks are dying relatively frequently due to the older age of the sample, so about 13 percent of subjects who were selected from the HRS sample died before we could contact or do an assessment on them.

Briefly, here are some of the characteristics of the ADAMS sample. You see the age, gender, and racial distribution here., The final unweighted distribution of diagnoses was about 36 percent diagnosed as normal, 28 percent CIND, and 36 percent demented. These are unweighted. Obviously you can't project to a national prevalence from these data, yet.

Characteristics of ADAMS Sample (N=856)

	N	(%)		N	(%)
Age			**Residence**		
70-74	171	(20)	Community	747	(87)
75-79	188	(22)	Nursing Home	109	(13)
80-84	224	(26)			
85-89	149	(17)			
90+	124	(15)			
Sex			**Respondent Type**		
Male	354	(41)	Self	657	(77)
Female	502	(59)	Proxy	199	(23)
Race / Ethnicity			**US Census Region**		
Hispanic	84	(10)	Northeast	120	(14)
Black, Non-Hisp	159	(19)	Midwest	167	(19)
White, Non-Hisp or Other	613	(72)	South Atlantic	238	(28)
			South Central	164	(19)
			West	167	(20)
Education			**ADAMS Diagnosis**		
< 12 years	442	(52)	Normal	307	(36)
12 years	195	(23)	CIND	242	(28)
≥ 12 years	218	(25)	Dementia	307	(36)

ADAMS Diagnoses (N=856)

Diagnosis	N	(%)
DEMENTIA (n=307, 36%)		
Alzheimer's Disease	228	(27)
Vascular Dementia	48	(6)
Subcortical Dementias	3	(<1)
Other Dementias	25	(3)
CIND (n=242, 28%)		
Mild-ambiguous	94	(11)
Mild Cognitive Impairment	4	(<1)
Cog Imp due to vascular disease	21	(2)
Stroke	34	(4)
Other Neurological conditions	10	(1)
Other Medical conditions	55	(6)
Depression	8	(1)
Psychiatric Disorder	2	(<1)
Low Baseline Intellect	8	(1)
Alcohol Abuse	6	(1)
NORMAL (n=307, 36%)	307	(36)

The diagnoses were distributed as shown on this slide.

Turning briefly to some preliminary data on the genetic information, the ApoE genotypes. Only 11 of the 856 people refused permission to do the buccal swab, and of those 845, I believe there was only one or two that technically couldn't be done from the sample. So we had a very high rate of successful completion of the buccal swab collection.

Apo E Genotype, by Diagnosis

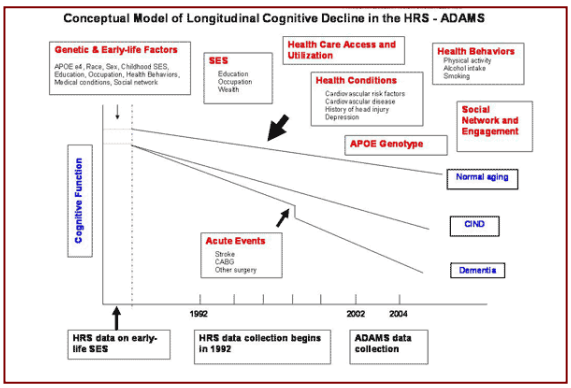
	Demented	CIND	Normal
	%*	%*	%*
ApoE e2/2	2.2	49.5	48.3
ApoE e2/3	11.7	25.5	62.8
ApoE e2/4	11.0	18.3	70.7
ApoE e3/3	10.0	24.1	65.9
ApoE e3/4	14.2	23.5	62.3
ApoE e4/4	41.3	26.3	32.4

*Weighted

The distribution of the ApoE genotypes is shown in this slide. This distribution is almost exactly the same as a couple of the other population samples that have collected this data, the Iowa Epese and the Framingham Study have collected the ApoE data, and it is literally almost exactly this same distribution, so we were happy about that. And we also have found in preliminary analyses, what we'd expect, that the ApoE e4 allele was associated with a much higher risk for dementia. Again, these are quite preliminary. About 41 percent of the ApoE e4/e4 individuals were demented compared to about 10 percent of the e3/e3 individuals.

JORDAN: Was the dementia in the 3/3 labeled as Alzheimer's?

LANGA: There were probably Alzheimer's disease diagnoses in here, but these results aren't broken down by diagnosis, yet. We haven't done that work yet.



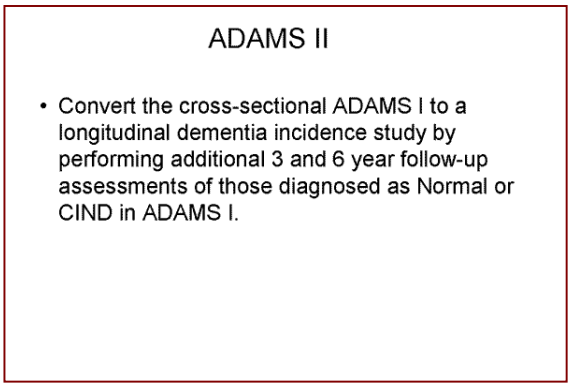
Just to finish, this is sort of a busy slide, my McArdle-like slide with longitudinal trajectories here. This is a "big-picture" conceptual model of how we think some of these data in the HRS and ADAMS might be used in the future to get at some of the issues we have been talking about over the last day or two.

Just to orient you: cognitive function is on the Y axis and time is on the X axis. In the HRS, we actually ask a number of questions about early life economic conditions, such as: Was your dad ever out of work for a significant amount of time? Did you ever have to move because of financial difficulties. We don't have in utero data, as we discussed earlier in the meeting, but maybe that's something we can think about for future data collection. It was interesting hearing in the presentation yesterday about how early life events might affect later cognitive function as well as cardiac function, and cardiac disease.

Again HRS data collection began in '92, so for the ADAMS sample we have up to ten years of prior information on what's been happening regarding health, cognition, economic status, and family life. The ADAMS data collection started here in 2001. Again all the complicated interactions we have been discussing regarding health in general, and cognition in particular, genetic and early life factors, SES, health care access and utilization, are shown in the upper part of the figure. I have been quite interested in how cardiovascular disease affects cognition, as well as other important health conditions such as head injury and depression. Health behaviors are also an important area of research as, observational data have shown that health behaviors (e.g., smoking, drinking alcohol, and physical activity) can affect cognitive health. Social networks and then genetics are also areas of interest.

This general conceptual model shows that as individuals reach middle age they have a cognitive reserve or their brains have developed in such a way that some have more reserve, more neuronal duplication, that allows them to start at a higher level of cognitive function; and then over time as they age there is some decrease in cognitive function. The individuals represented by the lower two lines have started at a lower level of cognitive function and have had different trajectories of decline with age. These are just hypothetical; there are an infinite number of trajectories that could result.

This person ended up in the CIND category when we saw them in ADAMS. Here's someone who ended up demented when we saw them. Acute events also likely have an effect. We have the ability to look at some of these acute events through the Medicare data that will be available to merge with the HRS data; events such as strokes, bypass surgery, any other surgeries.



Finally we're hoping to that the HRS funding is renewed this year, and we're hoping to continue the ADAMS process, an "ADAMS II," by converting the cross-sectional study into a longitudinal incidence study by following the individuals who were diagnosed as normal or CIND for two more in-home follow-up assessments at 3 and 6 years. so we'll have information about predictors of the incidence of CIND and dementia.

LINDAU: For a pinch hitter that was phenomenal. Thank you. I know I have questions.

(Applause.)

WEINSTEIN: So I think this is tremendously interesting, the fact that you have such a large sample is a great advantage, but certainly based on personal experience I think cognitive decline begins around age 16. And so I'm wondering why you chose to follow up the 70-year-olds and above rather than taking advantage of the full I guess age 55 and above, and that question is also motivated in part by the additional biomarker that you chose to collect; namely, blood pressure because one of the exciting parts of what you can do with these data as you move forward presumably is to try to tease out reciprocal relations between hypertension and cognitive decline as well as hypotension and the seeming relation between cognitive decline -- the effect of cognitive decline on blood pressure. So I guess this is a plug for catching more of your people earlier.

LANGA: That's a good question. There was a long debate and discussion about whether that would be feasible. David Weir can correct me if I'm wrong. I'd say if the budget constraint wasn't there, we would certainly be doing this. We actually developed some initial plans to follow younger folks. The general budgetary problem is that on a population basis, cognitive impairment is quite rare at age 50 to 60. It's definitely there, and it's starting; but in terms of finding enough cases to be able to make inferences and have the power to test hypotheses, the costs became prohibitive in terms of trying to follow some of the younger folks.

Again the second best way, we do have cognitive measures on these folks over time. Perhaps in the best of all possible worlds this imputation process will be able to impute back and have at least smaller confidence intervals about when someone does have dementia in the HRS. And other constituents have given us that advice also especially, as you were saying, in the cardiovascular issues and some minority populations, Hispanics and African-Americans, where these cardiovascular risk factors might be more common, it would be quite interesting to get an earlier look and see where the impairment is starting and see how the cardiovascular risk factors are interacting to cause both Alzheimer's disease and vascular dementia. But I'd say cost issues were paramount in not being able to look for cognitive impairment in 50 to 60 year olds. I don't know, David, if you wanted to add to that.

WEINSTEIN: I didn't think you guys had any budgetary problems.

WEIR: That's a false rumor.

DAVID: I was wondering what sort of consent you got to get the buccal swab and whether it's open to kind of any genetic testing in the future, or do you have to get further consent from people? That's an extremely rich potential source of future research, but it's also very controversial.

WEIR: Actually I was going to mention it in mine. Of the people who agreed to do ADAMS, virtually 100 percent gave the DNA sample. The protocol was that they surrendered that in perpetuity, signed off, so it's now being held in a repository at the University of Michigan pathology lab, and there's sort of a committee set up to review applications to do further analysis with it, so it is available for future work.

It was a cheek swab sample, so there's not infinite amounts of it. You can't immortalize the cells, so we're going to need in particular in the early years when techniques for amplification are weak and expensive, we're going to need to be very selective about that. We don't want to burn it all up with a bunch of gene snooping projects in the first few years, but certainly there's things that can be done as genes come on line.

LINDAU: I have two questions that are totally unrelated. One is about the literacy testing. Is anyone looking at those data in relation to cognitive function?

LANGA: Yes. Well, not officially yet. The data won't be officially available until the fall, but one of the collaborators at Duke, Guy Potter, is quite interested in this issue in terms of literacy and how it affects performance on the neuropsych tests and how it varies across geographic regions and racial groups and things like that. So nothing's been done yet, but there are a number of people with interest in that area.

LINDAU: I think it's great that you're doing that. The other thing I wonder is I did a study on literacy with younger women, reproductive-age women, and noticed that those who scored low on their literacy took a long time to sign their consent form, not to read the consent form, but actually to sign their name. We actually analyzed these data. It was very small numbers. There was a significant correlation if it took you longer than five seconds to sign your name, a young person, you were much more likely to have low literacy. I wonder if there are any handwritten tasks that would have been timed as part of the neuropsych testing?

LANGA: That's a good question. Well, there's a trail making test or trail making task which is timed, but that doesn't get at exactly what you are saying. Actually I think you and I have talked about our mutual interest in the mini mental state exam. One of the tests is to write a sentence, any sentence; and in clinic I'm always fascinated to see things like "my husband is a jerk" or --

LINDAU: "I don't like this test."

LANGA: "My doctor is a jerk" or something like that. So that written test is there. I don't think it's been timed at all.

LINDAU: We should look at those data. And a totally separate issue, but as an obstetrician, I know there's at least one other, looking for Bill in the room, and this relates to Maxine's comment about cognitive impairment starting at age 16; but one place we do see changes clinically in cognitive function is in pregnant women, and I have seen there are I think small investigations of this maybe even qualitative how do you think your memory has changed, but women definitely complain about changes in cognitive function during pregnancy; and it would be really interesting to think about whether the cognitive changes in pregnancy are in any way predictive of cognitive performance later in life or relate to cognitive trajectories for fetuses and what the mechanisms of that might be. So that's our next study. I can see several people in the room might be interested in the room in doing something like that.

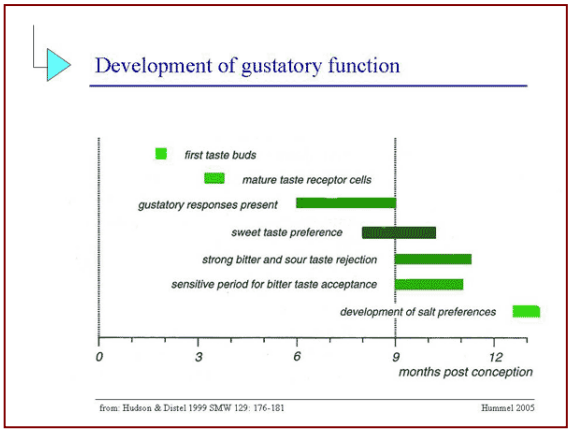
If there aren't any other questions, then we'll thank Ken Langa very much for an excellent talk.

Taste and Smell

Thomas Hummel

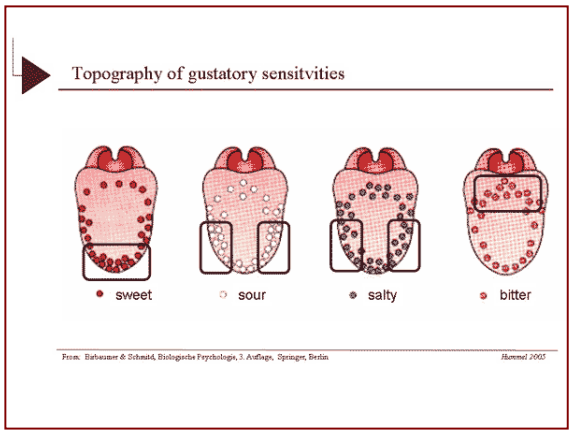
Thank you very much again for the invitation. It's my pleasure to be here, and also I have to say again I very much appreciate this collaboration, and I hope something will come from that.

Before I start I also wanted to say I've been assigned the task to talk to you about biomarkers. What I'm going to talk to you about today is taste. I will try to structure the talk into (1) basic characteristics of gustatory function, (2) different ways to assess gustatory function, the effects of sex and age; and (3) in the end I'll talk a little bit about interactions between smell and taste, about the different chemosensory channels that we have, and also I'll talk about taste function in smell loss.



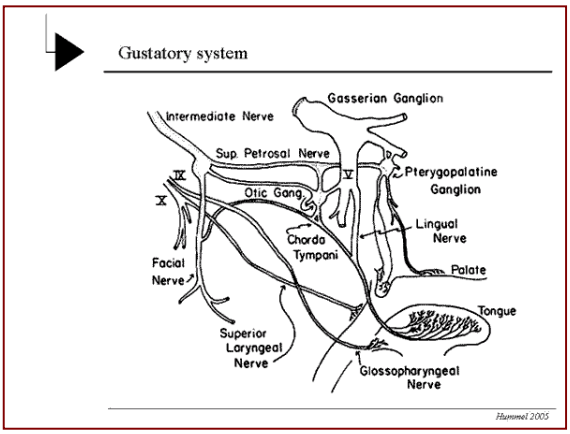
So to start with, when do we start to taste? Actually we start very early. We start in the uterus already to develop taste function. From month six on actually we are able to perceive tastes, and also there has been lots of speculation whether intrauterine exposition to tastants would shape the later experience one has with gustatory function or taste function later on in life and also as a consequence of that, whether this would affect dietary habits, or body weight, for example. But as I said before, very little study has been done on that. There has been much more study done on olfactory function intrauterinely, and there are also some very interesting studies that indicate that intrauterine exposure to certain odors actually shapes your preferences for certain odors.

There has been a very nice study done by a French group, by Benoist Schaal and coworkers, that investigated two groups of pregnant women. One group received during the last two weeks of pregnancy anise drinks and anise cookies, and the other group did not; they investigated the newborn babies of these two groups and guess what: The newborns from the anise group, they also had an anise preference. Similar effects have been shown very nicely in excremental animals, and lots of work has been done on that. Comparatively less study has been done on taste, so this is understudied in that sense.



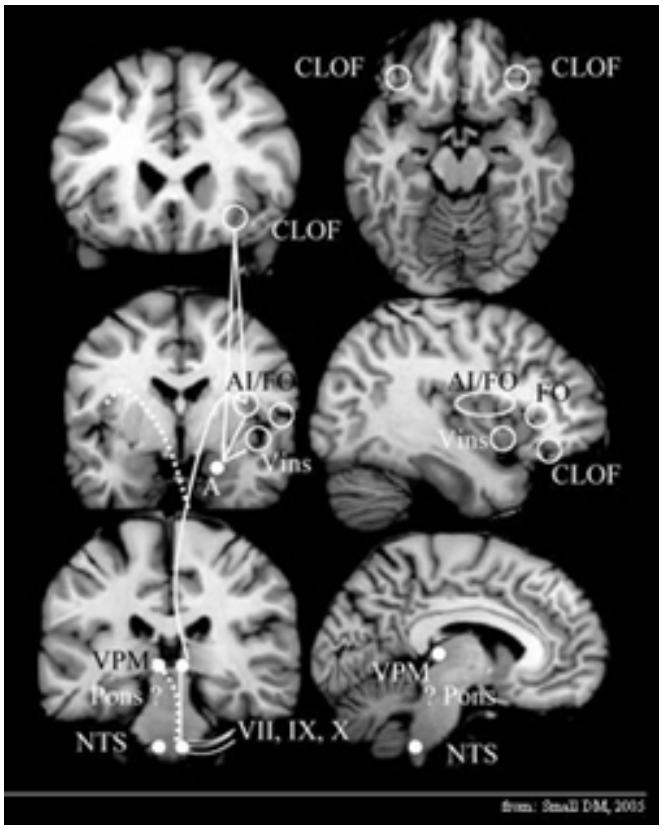
What do we know about taste? It starts on the periphery on the tongue, as you know; and on the tongue you have these taste buds, and we have different sensitivities on the tongue. Many of you may have learned that on the tip of the tongue we are most sensitive to sweet, and on the lateral side of the tongue we are most sensitive to sour and salty, and on the back of the tongue we are most sensitive to bitter. But also there has been transcription errors in some textbooks that indicate that we are only sensitive to sweet stimuli on the tip of the tongue or that we are only sensitive to

bitter stimuli on the back of the tongue. This is not true. Actually it's all over the tongue that we are sensitive to various stimuli (except the tongue center).



The tongue is extremely well innervated. There are actually three different nerves that contribute to taste function. It's the facial nerve, the glossopharyngeal nerve and the vagal nerve. So taste is extremely stable (compared to olfaction). If you are to lose the sense of smell, you only need to cut the olfactory nerves. To lose your gustatory sensitivity you have to cut six nerves, three nerves on each side. So its quite a stable system, and maybe this is one of the reasons why we also are not very sensitive to taste loss. If you section the chorda tympani (actually the facial nerve that provides gustatory sensitivity to the anterior tongue) in patients during surgery so that these patients after surgery do not have gustatory function on one side of the tongue, most of them would not notice that after surgery. This just comes to illustrate how insensitive we are to our chemosensory functions. I can cut your chorda tympani on one side, and so make your one side of your tongue actually blind to taste, and you probably would not notice.

Also, last year I think Johan Lundstrom talked to you about the difficulties people have knowing about their own olfactory sensitivity. People are very, very bad rating their own olfactory sensitivities. Similarly, that's also the reason why we have to measure gustatory function. Just asking people about their gustatory sensitivity or asking them about olfactory sensitivity is just not enough. We need to measure that to get a decent estimate of the chemosensory functions.



Gustatory system

NTS = nucleus tractus solitarius.

? = hypothetical pontine relay

VPM = vent post med ncl of thalamus

AI/FO = presumed primary gustatory area (anterior insula, frontal operculum)

Vins, FO = other cortical tastze areas (ventral insula, frontal operculum)

CLOF = caudolateral orbitofrontal cortex.

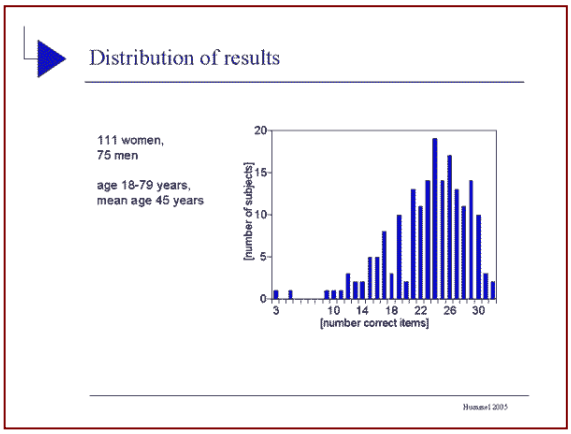
On a higher cortical level these sensations are in the thalamus, and much of this goes into the cortex in the frontal operculum. That's expressed here in this slide taken from Dana Small's work. What's interesting about taste is much of the information processing is going on in the orbito-frontal cortex, and this is also one of the characteristics that makes taste interesting as a biomarker because

this processing of gustatory information in the orbito-frontal cortex that relates taste very much to changes in cognitive function, so changes in cognitive function will also affect changes in gustatory function.

So how do we assess gustatory function? How do we assess tastes? Actually there's been a few suggestions; one is electrogustometry. This is like you apply a current on the tongue. It is like licking a battery, and it produces an "electric" sensation. That's been used extensively in the clinic, to not much avail, because there's very little correlation between responses to natural tastants and electrogustometric sensations. So electrogustometry is elegant, but it is actually not very useful.

The other thing is that you use tastants, you use solutions that are applied on the tongue; the downside the use of liquid solutions is messy. I mean you produce a solution of sugar; and you let it sit on your table for two days - and then it will be contaminated with bacterial growth. You don't want to put this sugar solution into the mouth of people any more. So, you have to renew solutions all the time, and this is extremely cumbersome; this is also among the simple reasons why taste is understudied. It keeps researchers from looking into that more deeply.

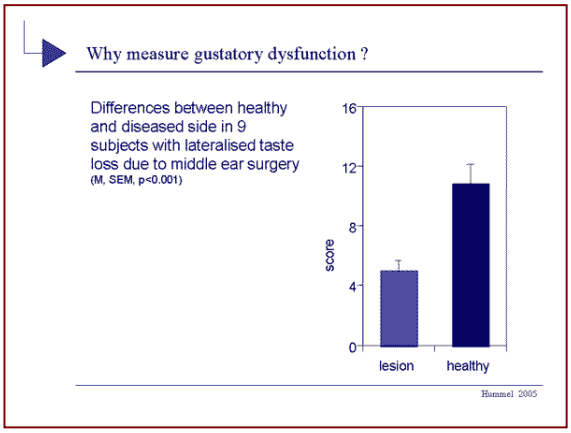
This was basically the reason why we came up with an idea to produce these "taste strips". These are strips of filter paper, impregnated with tastants; they are dried, and you can keep them on the shelf for years. As soon as you need one of those "taste strips", then you take those taste strips, place them on the tongue, and produce a gustatory sensation. The strips come in different concentrations of different tastants, so we would like to think it's a fairly elegant system to study taste.



There are normative data that are based on data from approximately 200 people. There are current studies on the way with this taste system, a number of studies look into therapy of gustatory dysfunction and what not.

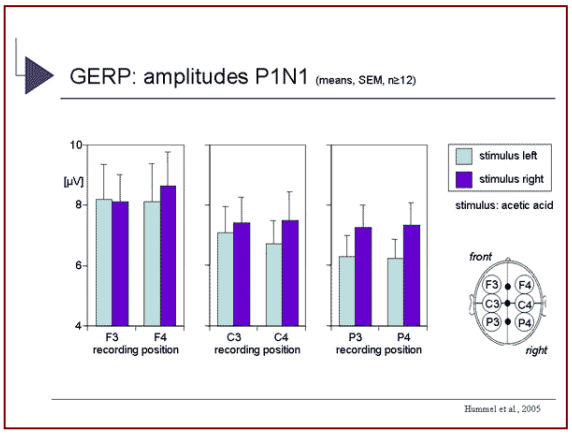
This is actually one slide that I wanted to show you which I referred to before. These are results from people who had a lesion of the chorda tympani. You see results for the healthy side, and this is for the lesion side. You see there's a tremendous decrease of gustatory function on the lesioned side. This goes absolutely unnoticed by the subject, so it's an interesting fact that one has to consider when one looks into taste.

Other ways to study gustatory function is to look at gustatory event-related potentials. The problem here is that we don't want to present liquids to the tongue because this produces a mechanical stimulation. We want to blow them on the tongue. Actually that's an apparatus that has been built by Zwaardemaker. Zwaardemaker is a Dutch physician who came up with this idea like 100 years ago and blew tastants on the tongue and produced gustatory sensations.



We use a somewhat more sophisticated equipment like this, but the idea is basically the same. We blow tastants on the tongue and produce gustatory sensations; this produces then these event-related potentials that are a measure of cortical function, of early cortical function. Using different stimuli, we are able to produce responses to sour, to sweet (actually chloroform is an

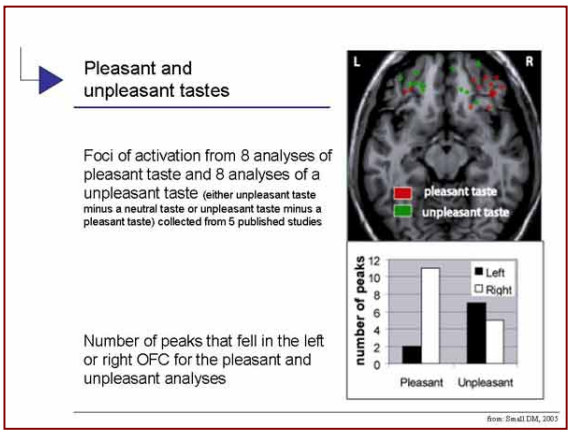
excellent sweet stimulus).



What does this tell us? This system let's us study gustatory processing; and what you can see here is when stimuli are applied to the right side, we get much larger amplitudes compared to left-sided stimulation which tells us that the right hemisphere is more involved in gustatory processing compared to the left hemisphere.

Another means to study gustatory function is functional imaging. This is again one of our stimulators outside the MRI chamber. What you see here are activation patterns in the brainstem following gustatory stimulation. These are data compiled from an article by Dana Small. What she shows here is that actually the orbito-frontal cortex is involved in gustatory processing. You can actually make up maps for the processing of pleasant and unpleasant tastes; and it seems like the medial lateral orbital frontal cortex is more involved in the processing of unpleasant taste compared to the lateral orbital frontal cortex.

This is very similar to olfactory function also. So pleasant odors are more processed here in the orbital frontal cortex on the more lateral side; and the unpleasant odors, they are more processed in the medial orbito-frontal cortex.



How about effect on age and sex? First of all, sex, what would be a guess who tastes better, men or women? It's the women. Women outperform men on all gustatory tasks that we and many others have studied so far. These are results for regional testing or whole-mouth testing. Red bars are women; blue bars are men. Men are hopelessly beyond the sensitivity of women.

Same shown here, different methods, blue are men; red are women. Again women outperform men for sweet, sour, salty, bitter. This also has been shown for umami. We also demonstrated this sex-related difference for event-related potentials.

Women respond faster to gustatory stimulation compared to men. The reason for this is unclear. There are some effects in the periphery, so women do seem to have more taste buds in the periphery, but why women are so much more sensitive to chemical stimuli, this is actually an enigma. I don't have a really good idea why this should be so, why should women smell or taste better than men.

LINDAU: Has this been studied in pregnancy?

HUMMEL: This has been studied in pregnancy.

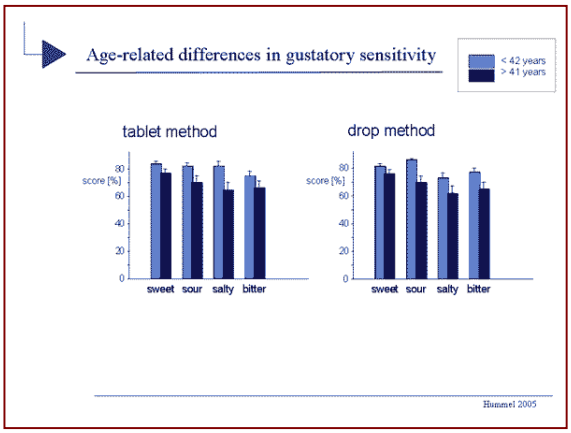
LINDAU: Does it increase?

HUMMEL: No, it doesn't. Actually sensitivity to "bitter" decreases during pregnancy, if you compare it to non-pregnant women.

There seem to be numerous cognitive changes going on in pregnancy, and that's probably also the reason why many women actually respond to odors during pregnancy so differently. This is not related to sensitivity per se. It seems to be the cognitive processing of these odors or tastants is different.

GROBMAN: Do those differences continue across all age categories? Like, for example, in older people, people past the age of menopause, things like that, those differences, do they still exist?

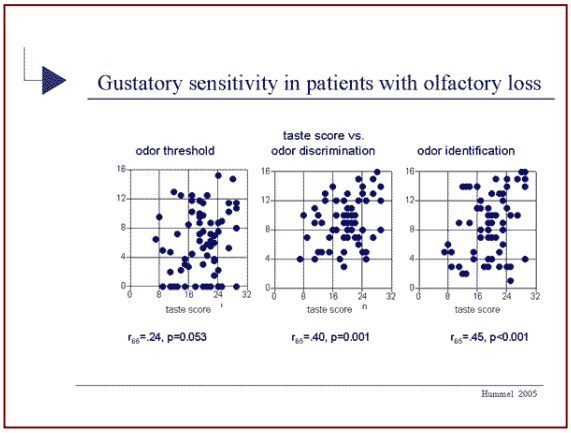
HUMMEL: What you see here are results in relation to aging. Results for women in red; men are in blue. You can see the differences are still present when women age, so when they are beyond menopause. This age effect is also present for different methods of assessment of taste function, and they are present all different stimulus categories, sweet, sour, salty and bitter.



That's shown here again. These are changes of gustatory function in relation to age. You see that there's a decrease with aging and also the variability of the results increases as people get older. So that may also be, as I said before, an indicator of cognitive function, but this is not really studied very well.

What makes "taste" complicated is that you're not just looking at a single system. You don't look just at taste. Actually the chemical senses are fairly complicated. If you take a cherry in your mouth, then you have this wonderful cherry experience; and this comes about by stimulation of the olfactory nerves mediating the cherry flavor, by the gustatory nerves mediating the cherry taste, like a little bit sweet and some sour to it. You also have mechanical activation. And you have trigeminal activation; and all this is put together by the brain, thankfully, into the cherry flavor. So if you have a cherry in your mouth, this is a multisensory interaction between many, many different channels, and it finally results in "cherry".

Actually what we did here is we put tastants into the people's mouth and looked at evoked potentials, how tastants would influence taste or olfactory function. What you see here is that when people have sugar in their mouth or if people have something sour, like citric acid in their mouth, then, you see, if they have sugar in the mouth and vanillin is presented to them, then the response latencies of event-related potentials are much shorter compared to the situation when there is something sour in the mouth. It's the other way around for CO2, which is pungent and also is very much associated with sour drinks. This is seen for all different components of the evoked potential. The interesting bit is that this happens extremely fast. So this interaction is already present at a latency of about 300 ms after stimulus presentation. So it happens on a subcortical level, these interactions between the olfactory and gustatory sensations.



Accordingly you could also expect that if people lose their sense of smell, there's also something happening. We studied that in patients with different degrees of olfactory loss. There is a significant difference between people with olfactory loss and people with no olfactory loss. So if you lose your sense of smell, you also lose this interaction between the olfactory and the gustatory system, and accordingly your gustatory sensitivity also goes down. So if one tries to study gustatory function, one also needs to take into account the olfactory system and also the cognitive situation of the subjects.

Again, the basis for this is this interaction that takes place between our senses between smell and taste here in the orbital frontal cortex, but there are also other areas where those two systems interact, namely in the insular cortex; studies in animals have shown that the thalamus is also involved in the interactions between smell and taste.

With this last slide I would like to acknowledge some of my coworkers who are all over the place, in Germany, many people in Sweden with whom we collaborate and also my friends in Dresden. Thank you very much for staying with me.

(Applause.)

JORDAN: Can you tell me a little bit about how you derive your score value?

HUMMEL: The score value is just a sum of correctly identified items, so, in a way, it's very crude.

NIELSEN: What's the relationship between this loss of olfactory or taste function and other cognitive measures? I mean does the decline in taste correlate with measures on specific neuropsychological tests, or how are you measuring or are you measuring this?

HUMMEL: We haven't studied that. It was Susan Shiffman from Charlotte, who did studies in early Alzheimer's disease about cognitive impairment; and in these patients she sees a decrease of gustatory function. There's also changes in gustatory memory, so there's a connection there.

LINDAU: Is there -- I'm sure there is -- literature looking at olfactory and gustatory function in relation to mental health or illness such as depression or cognitive impairment earlier in life?

HUMMEL: Earlier in life studies, that I'm not aware of, but there are changes in depression that you see for the olfactory system. So if you have a major depression, olfactory sensitivity goes down. There's nothing done in the gustatory system (as far as I know). The interesting bit is the olfactory sensitivity that comes back actually when you treat it with anti-depressants, when you successfully treat it. Then also your olfactory sensitivity bounces back to normal, so that has been shown fairly convincingly.

McDADE: What about stress? There's some suggestion that stress increases appetite for carbohydrates or fatty foods which could lead to different patterns of obesity, things like that. Any evidence that stress or affect could change different types of cravings or sensitivities to taste?

HUMMEL: Much has been done in olfaction, but with tastants I'm not aware of a study that looked into this. That's a biomarker that seems to be promising for many, many different studies, but it's hopelessly understudied. I think the major reason for this is that so far there has not been a decent testing system. So if you have a decent tool to assess that, then you probably also are going to answer these questions.

We would be extremely interested in what prevents taste loss as people age. Also the other question is what helps us to prevent olfactory loss when people age. This you have to think about. It's such a sad story. Every third one in this room, when we reach an age of 65, 70, every third one of us will be functionally anosmic. So we will have total olfactory loss. This means you can't enjoy your foods that much. You have to skip the nice Italian wines, it's a very sad story.

KUZAWA: Very interested why these taste strips that you have which you developed in thinking about some of the potential public health implications of using those, do you have some way to discern preference for fatty foods? I saw salty, I saw bitter, sour, and so forth. How do you get at preference for fat in food, or do you?

HUMMEL: This is only done by questionnaires. For fatty foods this is the only tool available. To test sensitivity to "fat" many people make custards very freshly. That's done by many different institutions - they use custards made freshly with different degrees of fat in there.

JORDAN: Is it true that smoking decreases your ability to taste, and did you take that into account when you studied your individuals?

HUMMEL: Smoking does affect your gustatory functions. Smoking also has an effect on olfactory function, but you have to smoke a lot. Actually there's studies that show that if you have like 20-pack years, so if you smoke one pack of cigarettes every day for 20 years, it's a lot, but this has basically no effect on your sense of smell. There's almost no difference compared to non-smokers.

If you ask people not to smoke actually one hour before they are being tested, then there's almost no discernible effect on olfactory sensitivity, and this is also similar for taste. You have to look very hard to see differences between smokers and nonsmokers in chemosensory sensitivity.

LINDAU: I'm going to take the liberty to ask one last question. What about the relationship between substance abuse dependents and olfactory or gustatory function? Has there been any work in that area?

HUMMEL: This is the only work that I know is with cocaine, and cocaine has some local effects on the mucosa, as you can imagine; but I'm not aware of other studies, no.

Detection of Late Life Dementia in the Community

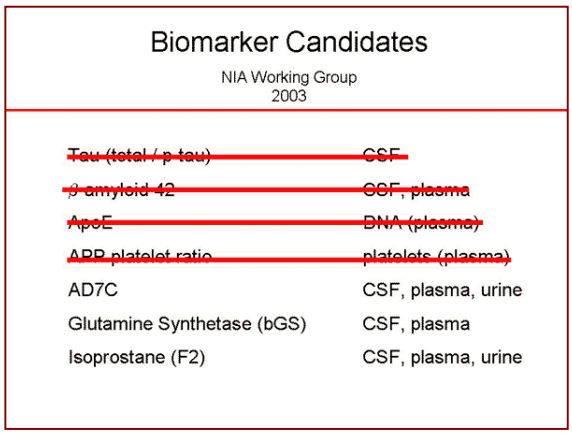
Christopher Clark

It's nice to be back, and I want to thank Stacy for inviting me to return.

I'm going to talk about detection of late life dementia in the community. My experience is not from a population-based survey platform but rather from working in a specialty clinic that has, as one of its research focuses, developing ways to detect the presence of pathologically specific changes in the brain as early in the course of their development as possible and to develop and evaluate biomarkers that are sensitive to treatment effects. Working in that environment I have access to high tech imaging and biochemistry methods as well as the ability to collect various biological samples, including spinal fluid, that would not be relevant to most community based population studies.

Where specialty clinic and population based researchers converge is in understanding that we must have validated biomarkers that can be used in community medical care settings if we are to make any meaningful advance in the diagnosis and treatment of the dementing illnesses of late-life. Assessment tools that can only be used in academic based specialty clinics will not produce meaningful progress in the use of biomarkers for dementia detection and treatment monitoring. In the ideal world we'd actually move dementia detection as close to the home as possible, reducing the need for physician time, and certainly reducing the need for dementia specialty clinics. At a minimum, it should be possible to do everything in the individual's own community; but that may be wildly optimistic. However, today I am going to focus on studies that operate at the level of the specialty clinic – community healthcare interface and present a little bit of our preliminary results.

So that said, what I want to begin by listing the current candidate biomarkers of late-life dementia and let you know how have they fared to date; an update of who made the initial biomarker team roster and who's been cut over the past couple of years and why. It sort of in part becomes a little bit of a talk in futility, this is who is left still standing because we haven't had any new biomarker recruits in the past couple years.



I'll also present initial results from our attempts to move the use of biomarkers for detection of late-life dementia into the community, and a little bit about the implications for the detection of neurodegenerative dementia in population-based studies. I'll leave it to you to figure out how reasonable that is.

What about the biomarker team? This is an edited list of the biomarker team as nominated by a 2003 NIA sponsored working group. The focus of this list, as is appropriate from a public health standpoint, is the detection of Alzheimer's disease.

There are a couple candidates that are not easily obtained in a home or community setting. These include those derived from CSF. This is unfortunate because CSF tau is the best and most reliable detector of Alzheimer's pathology in the brain.

The same can be said for CSF beta amyloid. The issue of beta amyloid detection in the plasma, while doable, remains quite controversial because the data is very difficult to interpret and because we really don't know the source (i.e. peripheral versus central) of plasma beta amyloid 42.

CSF tau, of course is the major protein that constitutes the neurofibrillary tangle, one of the two pathological lesions that define Alzheimer's disease in the brain. It leaches into the spinal fluid and can be measured with reasonable accuracy. Beta amyloid is the major component of the senile plaque, the brain lesions of Alzheimer's disease, and beta amyloid 42 simply defines the 42 amino acid form of amyloid which is the so-called bad amyloid or the amyloid that is thought to be most pathological.

Measurements of both in the CSF are very useful from a diagnostic standpoint, but not practicable in a community setting. And simply measuring plasma beta amyloid 42 is not a viable alternative because of the difficulty interpreting the implications of the results. Therefore, we can eliminate these two candidate biomarkers (CSF tau and both CSF and plasma beta amyloid) from the list of practical community compatible biomarkers of late-life dementia.

We heard a little bit about ApoE earlier today. I won't tell you any more except to emphasize that APOE e4 is a risk factor biomarker. It is not a diagnostic biomarker and doesn't help you out when you're trying to determine anything other than an APOE e4 carrier's risk for Alzheimer's disease compared to an individual who does not have an APOE e4 allele. It is not useful at the individual patient level. This takes it off the list of candidate diagnostic biomarkers.

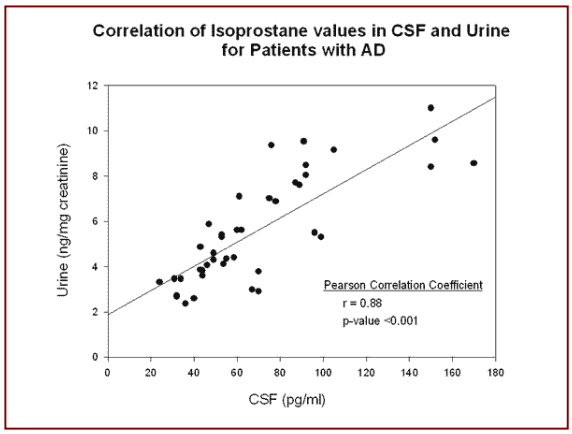
Amyloid precursor protein (APP) platelet ratio has been considered a candidate diagnostic biomarker. The ratio of the two common forms of this protein can be detected in platelets in an analogous manner to the detection of beta amyloid in the blood. There is some data to suggest this ratio is altered in patients who have Alzheimer's disease (AD) or who are at high risk for AD. So it remains a potential candidate, but the reports have only been coming out of one lab; and until it's been more widely validated, it will remain in the potential candidate category.

AD7C was proposed as a CSF biomarker of AD almost ten years ago in a series of intriguing papers. Then a couple of years ago the investigators reported it could also be detected in both the plasma and urine, which made it a potential community compatible marker for an Alzheimer's disease.

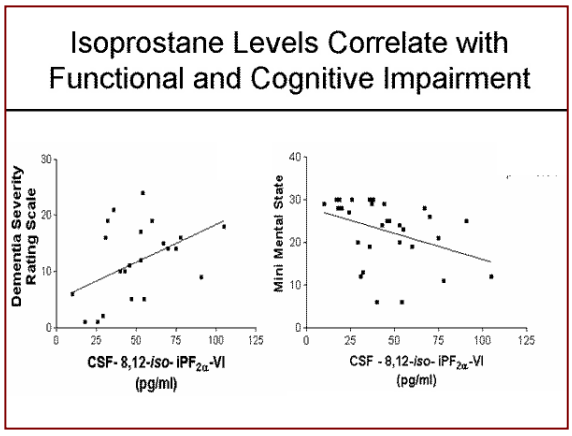
However, there was always an element of uncertainty because it is a relatively large protein (it is related to neuronal thread protein). From a biological standpoint it was never clear how this relatively large protein could get out of the brain, into the spinal fluid and then into the plasma and urine. So the academic community remained somewhat skeptical. And now it turns out that the AD7C protein may not even exist. This is based on the fact that it has not been possible to locate a gene that can encode it. In a paper that came out a couple of months ago, a team of investigators looked in humans and in chimpanzees for the gene sequence that was required in order to encode this protein based on the human genome data and could not find it. What they did find was the probably reason for the error. It appeared that the initial gene sequencing for the neuronal thread protein missed a couple of codons, including a stop codon. So they ended up with a very large protein that had a lot of common elements. The initial investigators on AD7C then developed an ELISA assay to detect the target protein which ended up reacting with a lot of proteins in the blood and in the urine and, therefore, gave a high positive response rate to this test.

So, it now it appears that not only does the protein not exist, but the gene that encodes it really is part of the junk DNA genome and has no known function. And so the authors of this most recent paper concluded, rather conservatively, with the statement that all previously published data related to AD7C needs to be reassessed. Thus, at this point, it is safe to say this biomarker has been kicked off the team.

So that leaves only F2-isoprostanes, which can be measured in CSF, plasma and urine, on the list of primary candidates. I talked a little bit about the potential for F2-isoprostanes in the earlier years in this conference. First just a little bit about why this is a viable biomarker candidate. Isoprostanes have a lot of characteristics that you would like in a community biomarker. They're chemically stable end product of lipid peroxidation. They're not subject to any further enzymatic metabolic or catabolic changes once they are formed in-vivo, and they actually pretty stable in the collected CSF, blood and urine. You can stick them in the refrigerator for days and you can even put them on the shelf at room temperature for at least 24 hours without degradation. They are linked to Alzheimer's disease in that they are elevated in the areas of the brain that have plaque and tangle pathology disease and not elevated in another common neurodegenerative illness (frontal temporal dementia), schizophrenia or control brains. We have focused on the 8,12,-iso F2 alpha IV isoprostanes. Isoprostane levels measured in the spinal fluid have a very good correlation with levels in both blood and urine as demonstrated in this graph correlating CSF and urine values in patients with Alzheimer' disease. The urine is expressed as nanograms of isoprostanes per milligram of creatinine in order to standardize it.



Interestingly, the levels in patients with dementing illnesses appear to correlate with the severity of the dementia. In this respect it differs from CSF tau, which does not seem to increase as the severity of the disease increases. This characteristic has pluses and minuses associated with it. Because it's elevated with disease severity, it could potentially serve as a marker for disease activity and, as such, be a marker of treatment efficacy. However because it's low in the early stages of the disease, when the symptoms are mild, it is not a good candidate biomarker for early detection of Alzheimer's disease during the prodromal state phase. The following two graphs show the relationship between urine isoprostanes values and clinical markers of functional impairment (the Dementia Severity Rating Scale, whose score goes up with increasing severity) and cognitive impairment (the Mini Mental State, whose score goes down with increasing severity).



This finding was demonstrated in a second study carried out by Dr. Mony deLeon at NYU. All the participants were nondemented when they entered this pilot study. They were followed with serial cognitive testing and spinal taps every year. Over time most who progressed to mild cognitive impairment and Alzheimer's disease had isoprostane levels that increased with each evaluation, lending additional support that it is a biomarker in CSF that increases with progression of the disease. And based on our earlier findings of a good correlation between CSF and urine values, the same changes found in the CSF in the deLeon study, should be seen in the urine.

So what about the initial results in the community health care screening program? The basic question is: Can you detect late life dementia, (mostly Alzheimer's disease), in the community using brief clinical assessments obtained in conjunction with a biological biomarker of pathological changes in the brain? So our goal is relatively simple. We simply wanted to see, in a community setting, if we could successfully categorize

individuals into one of four groups: those with successful brain aging; those with mild impairment; those who were demented, with AD; and those who were demented with something other than AD.

Approach

Test: in the community

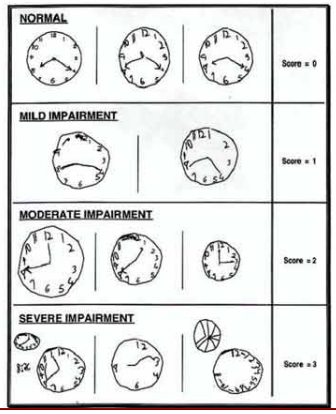
- Brief screening assessments
  - Cognitive status: direct testing
  - Functional status: observation of family member
- Biological marker: urine F2 isoprostanes

Validate: Memory Disorders Clinic

We used a very brief cognitive screening measure to take a direct measure of their cognitive abilities with a test that took, on average, 8 minutes to administer. In addition, to get some idea of their functional status we asked a family member (most often their spouse or an adult child) to complete a brief, 5 item, multiple choice questionnaire about their memory and ability to carry out routine tasks, with an emphasis on noting changes that have occurred over the past year of two, as it is changes in functional status that are the most sensitive indicator of the neurodegenerative dementias.

To link this information to a biological marker; we collected spot urine samples for F2 isoprostane measurements. The clinical category assigned on the basis of this data was then validated by a full assessment in our Alzheimer's Disease Center Memory Disorders Clinic.

Here's a brief description of our seven-minute cognitive screening test and the five item functional questionnaire. The latter was extracted from an 11 item multiple choice functional questionnaire that we have been using for almost 15 years. Currently the questionnaire is paper and pencil based, but the goal is to develop a web based version that could be completed in any community setting, whether it be in a health care clinic, community library or at home.



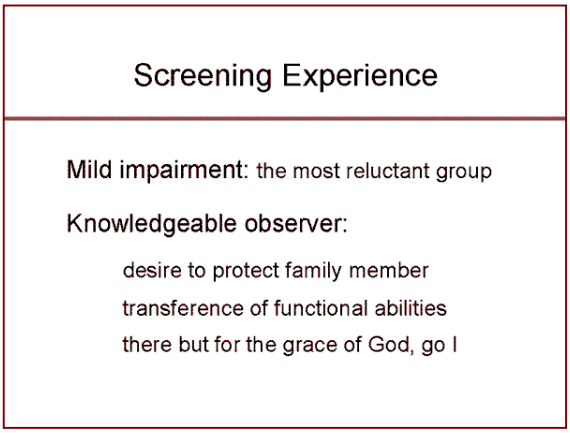
The brief cognitive screening touches on all the major cognitive elements that are affected in neurodegenerative dementing illnesses, including question to assess orientation and memory, (three words plus their own 7 digit phone number), and a serial subtraction test, (serial threes). Serial subtractions are only done with patients who have 12 or more years of education; and since we're counting error scores, dropping it for those with less than 12 years of schooling has little impact on the final score. There is a verbal fluency test (animal naming) and lastly, they are asked to draw a clock, which is scored from 0 to 4 based on how it compares to a set of standard clocks. The tester then simply adds up an impairment score.

Clinical Category

Impairment score	Clinical Category
0 - 4	no cognitive impairment
5 - 8	mild cognitive impairment
≥9	probable dementia

Thus, the cognitive domains assessed; constructional praxis, language, memory and orientation, are the major ones affected in demented illness. The screening test can be administered by anybody with a high school education and the appropriate personality. The tester simply adds up with errors, and that servers as the subject's impairment score. The lower the score the better. Zero to four represents no clinically meaningful cognitive impairment, five to eight is mild cognitive impairment and nine or higher represents probable dementia.

The functional questionnaire basically looks at their ability to perform tasks that relate to their memory, their language, their orientation, their ability to make decisions and their ability to navigate in space. This is a multi-choice questionnaire, consisting of five questions. For each of the five questions, Individuals circle the answer that most reliably reflects the current status of the person on whom they are reporting. The person completing the questionnaire is usually the spouse or an adult child, Adding up the numbers circled in response to each question (they are in order of severity) provides the numeric score.

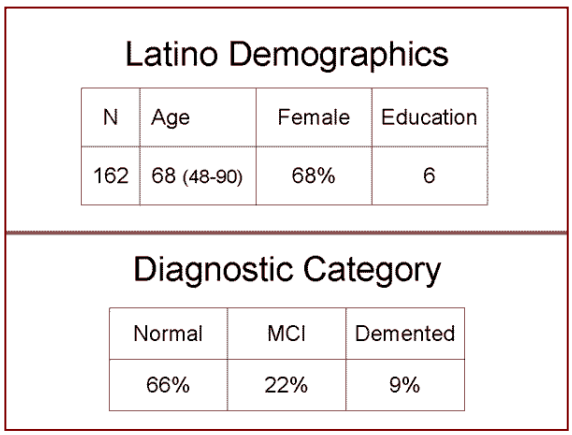


How has this worked in the community? Fairly well, although, not surprisingly, individuals who score in the mild cognitive impairment range on the cognitive screen and/or functional assessment questionnaire turn out to be the most reluctant to have fully (validating) evaluations done in the Memory Disorders Clinic. About 10% of individuals who score in the mild impairment range decline to have a full evaluation done. While I am not certain of the reason for this reluctance, I think there's a natural tendency, particularly among spouses, to protect the family member and to avoid facing any problems that may actually be going on. In addition, the spouse may have a tendency to under rate the functional questionnaire as they transfer their own abilities to the subject and, therefore, under rate any impairments they may have, producing a disconnect with the subjects performance on the cognitive screening test.

Here are the results from the first 92 subjects evaluated in the University of Pennsylvania West Philadelphia community Geriatric Practice. The practice is more than 50% women with a substantial minority component. They are relatively well educated, as noted in this demographic slide.

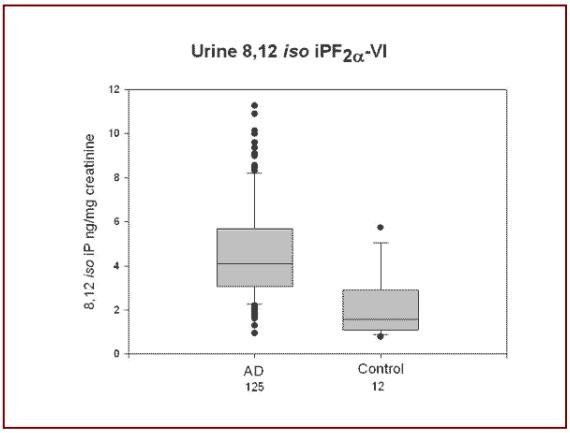
When we take a look at the overall score, the cognitive assessment results in conjunction with the functional scoring, it turns out that about 55 percent of the patients score within the normal range. The mild cognitive impairment range is about 30 percent, and about 14 percent of them fall within the demented range. And about half have had gold standard cognitive evaluations. Following the more extensive Memory Disorders Clinic evaluation a few patients shifted diagnostic groups, a couple moving from mild cognitive impairment to normal and some from mild cognitive impairment to demented.

We have also used a Spanish version in a Latino community health center in North Central Philadelphia. The Latino group has a younger age in general because this was not a geriatric focused clinic and we approached individuals attending the clinic who were 50 years of age or older. Their average education is significantly lower than the West Philadelphia Geriatric Clinic. However, as with the West Philadelphia patients, over half fall into the normal range on the cognitive and functional screening. Almost a quarter fell into the MCI group and 9% scored in the demented range.



In a separate exercise, we administered the screening battery to individuals already assessed in the Memory Disorders Clinic to see how well it performed when we were certain of the diagnostic category, and were pleased to see that, in fact, it seems to perform relatively well. In the control group the average score on the cognitive screening was 1, with a range of zero to 3 and the standard deviation of 1.3; and for the functional

assessment questionnaire, it was close to zero with a range of zero to two and a standard deviation of 0.5. The averages and ranges for the other two groups are presented on this slide.



What about the urine isoprostanes? Here are the results of the sample that we have both screening data and full Memory Disorders Clinic evaluations. You can see that while there's a reasonably good discrimination for urine 8,12 F2 isoprostanes between Alzheimer's disease and controls, there is an overlap and a few outliers in the control group. So this test is not going to be the only biomarker test you have to do. And it should not be interpreted in the absence of some assessment of cognitive and functional impairment.

That's been the history or the experience of all biomarkers in neurodegenerative disease. The general rule of thumb is the greater the variety of biomarkers you collect, the more informative they're going to be and the more accurate your ultimate diagnosis is going to be.

So regarding the question of can you detect late life dementia in the community; I would simply note Eleanor of Aquitaine's retort to her husband Henry II, in her attempt to convince him that she could block his annulment of their marriage. To quote Eleanor: "It's possible". Henry was skeptical. However she was correct.

Nevertheless, detection of late life dementia in the community is always going to require some degree of a clinical assessment in combination with an easily obtained biomarker. I don't think you can get away with using only biomarkers. Using brief cognitive screens alone it should be possible to categorize individuals with some degree of reliability, into cognitively normal, definitely demented, and those with mild cognitive impairment. It is this latter category that will prove the most difficult to know how to proceed. Using a biological marker will certainly help, but you really need both, and you need to interpret each in relationship to the other. So that's the update from the clinical perspective biomarkers.

(Applause.)

LINDAU: Are there any questions?

LANGA: Thanks, Chris. That was a really helpful review. I have a question about how you're going to interpret the isoprostane data. So it seems actually both clinically and from a research perspective the key sort of economic question or cost effectiveness question is how much additional bang for the buck the biomarker gives you over the prescreening exam so how much more precision does that give you per dollar, and are you going to analyze the data that way or present clinicians with just the screening data and see if that changes things at all?

CLARK: The answer to that is that it really depends on several things; the context in which you are using it; what's its public health implications are; how well the preliminary results hold up when tested in multiple communities; and can an assay be developed that is cheap enough and reliable enough to be run in a non-for-profit community hospital laboratory.

The data that I presented is based on assays that were HPLC tandem mass spectroscopy. That's expensive laboratory technology that is unlikely to be generally available in most communities. It may require the development of a reliable ELISA assay for it to be widely available.

The second question regards the utility of doing this. Right now the utility isn't that great. Although I'm a firm believer that it's great to know if the early changes that a family member observes in their spouse or parent are the beginnings of dementia because all of the planning that needs to take place; estate planning, long-term care planning, establishing a durable power of attorney, etc, but those issues alone are often inadequate to motivate people to seek an early evaluation of mild memory problems.

However, if a truly effective treatment for Alzheimer's disease becomes available and acceptable toxicity, such as the successful development of a modified vaccine that can clear amyloid lesions from the brain, individuals might be willing to accept significant side-effect risks to halt the progression of lethal dementing illness such as Alzheimer's disease. For example, if you told me I had a 17 percent chance of getting encephalitis and about a 3 percent chance of dying and a 90 to100 percent chance of avoiding becoming demented, would I take the vaccine? I would say that I would be willing to take that risk only if you could guarantee that I have the disease and to accomplish that that availability of a reliable panel of biomarkers including the isoprostanes would probably be quite valuable and certainly better than having a brain biopsy. Of course that answer is predicated on the assumption that there will eventually be pathologically targeted therapy that is truly effective and that it will not be risk free.

On the other hand, if the treatment for Alzheimer's disease were the equivalent of fluoride in the water for the prevention of dental cavities, there would be no need to develop reliable biomarkers. Everybody would get the treatment and from both an individual and public health standpoint, we would not need to only target those who are destined to get the disease.

LINDAU: It will be interesting. We are wanting to have a small break before the next session so we should probably stop, but thank you very much, Dr. Clark.

CULTURAL AND RACIAL/ETHNIC CONSIDERATIONS IN BIOMARKER COLLECTION

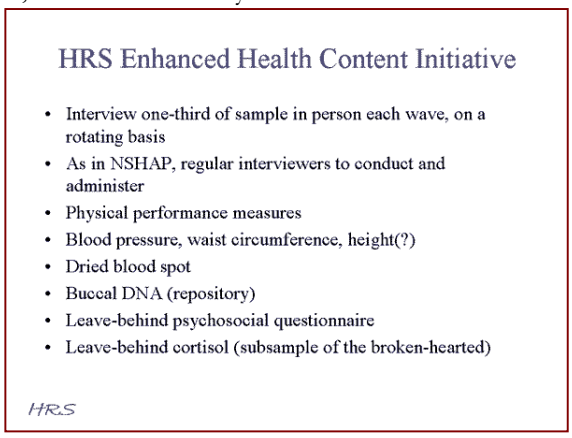
The Health and Retirement Study

David Weir

The HRS is a nationally representative longitudinal study. We interview about 20,000 people every two years. The primary mode of interview traditionally has been telephone, which sets some pretty big limits on what we are able to do. We do over the telephone some direct cognitive assessment, using relatively well-known assessment measures that were discussed previously by Ken Langa.

Because there's no telephone-administered biomarker tests, for us to do anything in this area we have to break out of our traditional modes. We have done some of that in supplemental data collections, and I will talk a little bit about those. We're also proposing going forward with a substantial new initiative moving into this area so before I get into the details on what we've done and focusing on minority response rate issues, let me just talk a little bit about those future attractions.

The design is to interview one-third of our sample in person each wave on a rotating basis; and if that kept going, then everybody would get biomarkers every sixth year, and so each wave you'd have a one-third sample who got the biomarker measures. As in the NSHAP study, regular interviewers will be used to conduct and administer what we do. We are not at this time contemplating hiring nurses or other methods for collecting specimens. Among the things we want to do are physical performance measures, and I'll show you some of our experience with those in the past wave. Those will include the things that we did already: grip strength, lung function test and timed walk, and we're contemplating adding perhaps balance measures to that. What we have not done before is blood pressure measurements and waist circumference. We did height on a small experimental group, and we may continue to do that.



We will do some dried blood spots. Exactly what assays will be done with blood spots will reflect what technologies are available. Currently we plan to do cholesterol, hemoglobin A1c, and C-reactive protein. We intend to ask for DNA samples in a mouthwash buccal collection, and Ken mentioned that we did this with a standard cheek swab in our ADAMS supplement. The proposal asks only to collect and store. We aren't proposing as part of the core HRS, which is really a data collection study, to do any specific analyses; but they will be then available for qualified research purposes.

We will also have as a leave-behind questionnaire some self-administered content on psychosocial, social support and a variety of other domains that we're in the process of narrowing down. We also experimented with that last year.

For the rest of today's talk I want to focus on the theme of this section which is special concerns in dealing with minority populations in the United States, and so I'll talk about our experience with a number of different kinds of data collection, our core survey, some of the supplements that have tried to get at biomarkers, and our physical measures in 2004.

Basically the core HRS over all shows essentially no differential in response by minority identity. There is a small difference in the composition of non-response: whites are a little more likely to request permanent removal from the study, but a little more likely to respond assuming they are still in. Those two things cancel leaving the overall response rates very similar across groups.

We also in 2004 conducted a screen to get a new cohort, people who were born between 1948 and 1953, roughly 51 to 56 years old in 2004. That involved two stages, first knocking on doors in randomly selected areas across the country to identify people who are in this age group, and then asking the target group to join the study. Of course in the first stage we found a lot of people who were older or younger than the ages we were looking for. For those of you who are familiar with the NSHAP study, the people we found who were in the older age groups already represented in HRS became the sample base from which the NSHAP study is being drawn.

HRS 2004 New Cohort Screening and Interview Response Rates, by Sample Domain (Percent Minority)

	Non-Minority	High A-A	High Hisp	High Both	All
Screener	91.9	91.8	92.0	88.4	91.2
Interview	77.5	78.1	77.3	83.5	78.4
Total	71.2	71.7	71.1	73.8	71.5

HRS

The screening response rates in the top row are categorized by the minority composition of the area. We oversampled areas with high densities of African-Americans or Hispanics or both, and that's the way we intended to generate an oversample in our net cohort of those minority groups. The response rate to the screener was around 92 percent in all areas except that the area that had high density of both minority groups was a little bit lower for an overall rate of just over 91 percent. I've listed that as though that were the response rate of the cohort we were trying to recruit. Of course if people didn't respond to the screener, you don't know how old they are so you don't really know what the exact response rate is by group; but the closer you get to 100 obviously the more sure you are that it's close to that for everyone.

The next step for us was to ask the people in our age group of interest to do the interview, and there we compensated for the lower screener response rate in the double high density area with a higher interview response rate, so the combination was around 71 or 72 percent across the board. That's a little lower than we've done in the past, and that's quite consistent with what's happening in surveys across the country, particularly at the screener level. It's getting harder to get people to respond to knocks on the door. But the net result of course is then that there was no substantial differential in response rates by minority location.

Minority Oversample Rates in HRS 2004 New Cohort

Table C32: Race and ethnic distribution of CPS and HRS samples of respondents born 1948-1953

	CPS	HRS	Ratio HRS/CPS	Ratio minority/ White, other
White/other	80.9%	67.9%	0.84	1.00
Non-hispanic black	10.6%	17.4%	1.64	1.96
Hispanic	8.5%	14.7%	1.73	2.06

HRS

Then we did a further check. We had targeted a 2 to 1 oversample rate. That is, we wanted to have blacks represented and Hispanics represented in our sample at double the rate of whites in the population. What you see is that based on the CPS you would have about 10 percent black and 9 percent Hispanic with no oversampling. In our sample it's 17 and 15 percent for the minority groups, and only 69 percent instead of 81 percent for the whites and others. When you take those ratios, we came out almost exactly on target with our 2 to 1 oversamples. To sum up: on the core survey whether it's telephone or in person, there is little or no differential in response rates for minority groups.

Now I'm going to talk about some of our supplemental efforts. First I'm going to talk about some things that except for the ADAMS did not involve biomarkers. The top panel there are mail surveys. These are things where we send them a questionnaire in the mail typically with a $20 check in advance for them to fill this out and send it back.

The HUMS is a human capital survey. That was asking people about expenditures on college for their children. CAMS is now a panel which is the consumption and time use mail survey, and that was done in 2001 and then repeated in 2003 and will be repeated again this year. Typically we get 80 percent or so response rate to our mail surveys done this way, and we're pretty happy with that performance. What you see over at the right, and this was taken from our renewal proposal where we standardize a number of our statistical procedures, what we did was we ran probit regressions, (using STATA's dprobit procedure), and we can report coefficients that are the differential response rate at the mean of the variables that you're controlling for, and what's controlled is education, age and gender.

Response Rates and Minority Differential Response Rates to Supplemental Data Collections in HRS

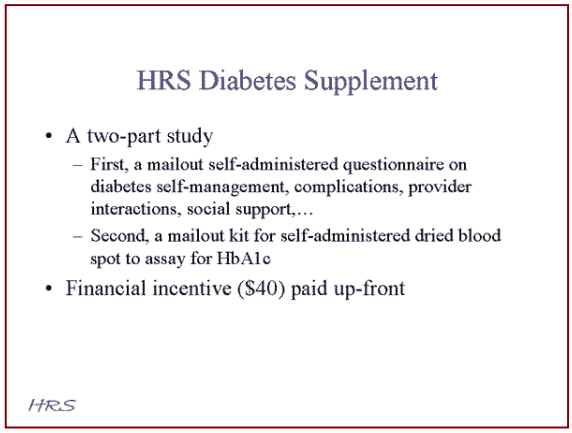
Supplement	Number of Responses	Response Rate	Probit Results[b] Black	Hispanic
1999 Experimental Mailout	2454	85%	-10.0	-5.4
2001 HUMS	3040	81%	-15.9	-11.7
2001 CAMS	3866	80%	-15.3	-13.2
2003 CAMS	3254	80%	-18.7	-17.1
2003 Internet Study	2176	59%[c]	-19.3	-13.2
ADAMS	856	56%	6.5	7.2

HRS

So these effects are the difference in the average response rate controlling for education, age and gender between blacks, Hispanics and the residual white and other category. What you see in all those self-administered questionnaires and including the Internet survey down at the bottom is a 15 percent or better penalty for minorities. So we do not get quite the participation in these self-administered surveys of minorities that we get in the core survey that involves a person calling up and conducting the interview.

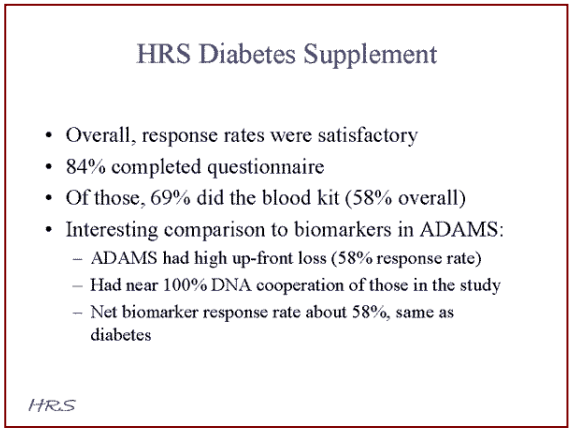
The ADAMS study, which was a three-hour in-person visit, on the other hand, actually had slightly positive coefficients for minority participation. Those aren't statistically significant, but certainly there was no minority penalty in the ADAMS study; but of course they had a much lower overall response rate. If we summarize this, we find that in contrast to the core survey, on self-administered activities there's lower response rates from minority groups. The question then is what happens if we do things that involve biomarkers?

The main test we have of this is a diabetes survey that we conducted, and let's just say for Maxine's benefit this was done on the cheap. Compared to the Cadillac of ADAMS, this was a beat-up Volkswagen of a study, and we were trying to see what we could get in this mode. It combined two parts. First was a mail-out questionnaire just like the ones I showed you the results for except that the content was about diabetes, about their self-management, about complications they've experienced, their interactions with their providers, the amount of social support they have and a number of things that are fairly commonly used in analyses of diabetes.

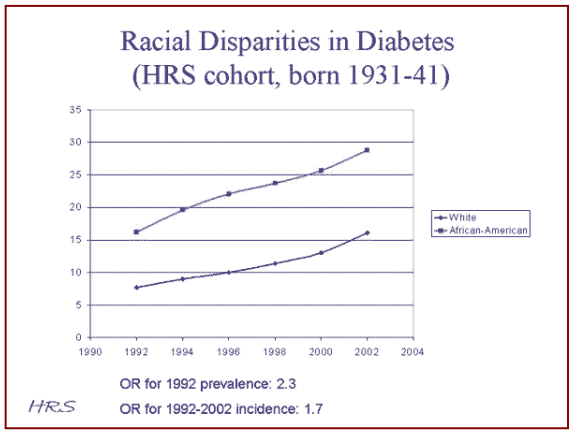


The second stage then was a test kit for a dried blood spot test for hemoglobin A1C which I'm sure most of you now know is a measure of the average level of blood glucose over two to three months prior to the administration. It is something that does not require any fasting. It can be done any time of day, and so the availability of such a test for a critically important clinical marker for diabetes and the quality of management of diabetes in a dried blood spot made this study really compelling. Because it was in two parts we gave them a $40 incentive rather than a $20 incentive but gave it all up front and mailed them in two stages.

Overall the response rates to us were satisfactory; 84 percent completed the questionnaire, and of those almost 70 percent did the blood kit for a net of about 58 percent. Now, it's interesting to compare that to the ADAMS DNA biomarker. In ADAMS you had a very high up-front nonresponse. Only 58 percent agreed to this fairly demanding in-home assessment. But once you're in there with this group of people who are willing to do that, most of them were willing to do the biomarker.

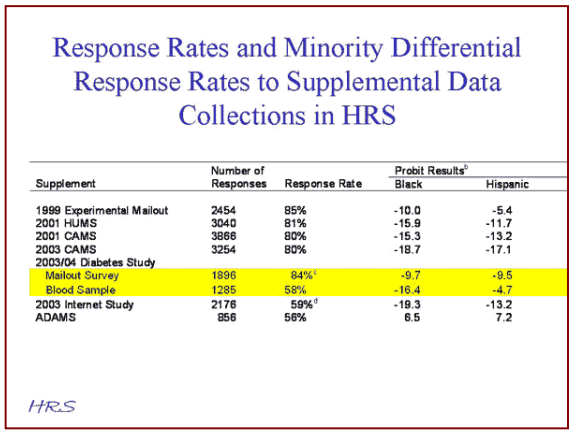


The diabetes mail survey, on the other hand, had relatively low up-front nonresponse but then a much bigger loss among people who had done the mail survey who wouldn't do the biomarker. So we ended up with actually the same net response rate for the biomarker.



Part of the motivation of course for looking at diabetes is the importance of racial disparities in diabetes in the U.S. population. This is taken from the HRS original cohort, the 1931 to '41 birth cohort first interviewed in 1992 and then followed over time. First we compare the prevalence rates of diabetes in 1992 when they joined the study. The odds ratio for prevalence in 1992 was over 2: African-Americans were twice as likely to report a diagnosis of diabetes in 1992 as whites the same age. Next we follow the groups longitudinally. The odds ratio for incidence, that is the probability of getting diabetes if you didn't have it in 1992, was again close to 2 for blacks versus whites.

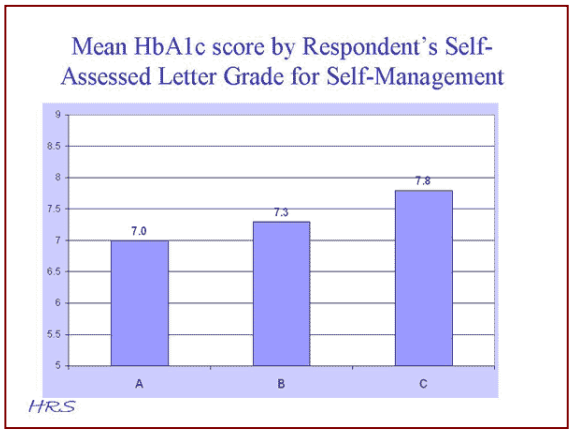
Now let's look at the response rates and how the minority differentials compares. This now fills out that earlier table with the results from the diabetes study. What you see is we had about the same or maybe a little less racial penalty, minority penalty, on the mail part of the questionnaire. The differential was around 10 percent versus the 15 percent on a lot of the others, so similar kind of results.



Then we can ask of the people who did the mail survey what's the penalty on doing the blood sample, and there particularly the African-American participation rates were just really low, and so this I think is a concern. Of course there's a lot out there about the history of inappropriate experimentation on African-American populations making them distrustful, particularly distrustful on things involving blood; and this would seem to be in line with that. Certainly I think it suggests that studies that want to collect biomarkers in population surveys of minorities need to pay some attention to how to encourage response and cooperation.

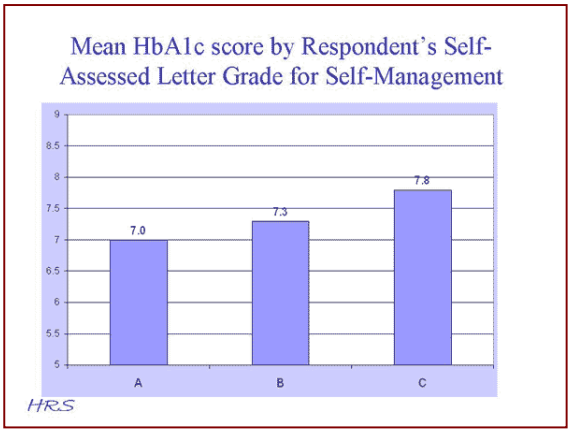
SMITH: Can I ask a question? The last feature of the second design, the one where you get in the door and then you suffer a loss second is that you collect some very helpful data to analyze nonresponse. I was just wondering if you had any observations in terms of the characteristics of those folks who responded to the initial survey but then dropped out of the biomarker collection?

WEIR: We've only done a little, and they're not surprising. I mean there's a little bit of an education differential, a little bit of an SES differential more generally. We in theory have somewhere some qualitative reports from a callback effort to find out what was happening. We haven't had time to digest the qualitative information into anything quantitative.



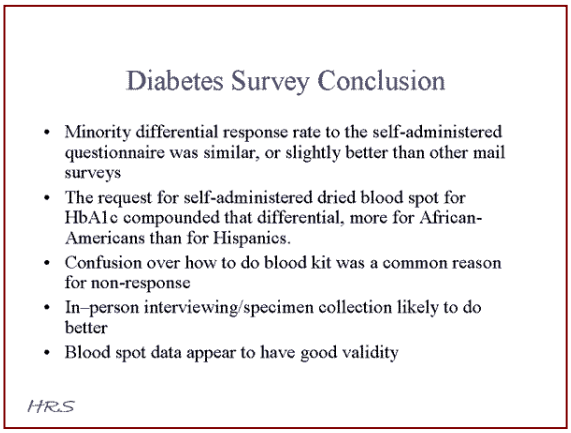
The one thing that came out is that confusion about how to do the blood test was a factor, possibly more for Hispanics than for blacks, but again we don't have the quantification to say that for certain. I do think one of the issues is you open this box up and there's these instructions in this thing and they're confusing, so you just put them over here instead. One of the conclusions that I will draw from our mail survey experience is that doing biomarker collection in person is likely to be more successful than asking people to do it unaided, unguided on their own It's cheaper if you can do it by mail, but you may pay a pretty stiff price in non-response.

Let me turn to some evidence on the validity of these biomarker data that we collected by mail. First, we look at the mean hemoglobin A1C score by medication regime. Just as you'd expect we find that people on insulin have high values, people on oral medications only, lower, and people who were not being treated with any medication had average numbers well below 7.

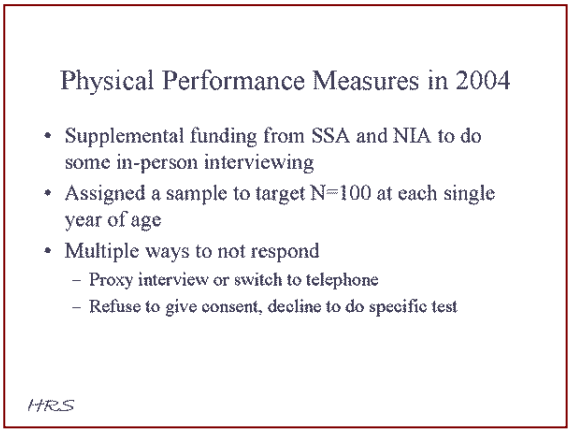


One of the questions from the diabetes survey that's fairly easy to use quickly is we ask people what grade would you give yourself for how well you manage your diabetes, and this correlates with things like how often they say they missed their medications and some other things. And you see that the people who say they give themselves an A had the lowest (best) A1C scores, and the people who gave themselves a C or worse had the worst. So there's in some sense a double validation going on here in that people do know whether they're doing a good job or not, and that does correlate with their outcomes.

Finally the differentials by race and ethnicity in A1c scores do to some extent mirror the differentials in prevalence. What I haven't done is try to adjust this with duration of the disease or other things. The fraction of African-Americans who are actually already progressed to insulin is quite high, for example. So one of the points to take from this is despite the fact that response rates over all were considerably lower for the minority groups than they were for the whites we're still able to capture a racial difference in the A1C score. The non-response does not appear to have been strongly related to severity of the disease, for example, in those populations.



All right. So conclusion with the diabetes study, the minority differential response rate to the questionnaire was similar or better than other mail surveys, but the request for a self-administered dried blood spot compounded the racial differential and more for African-Americans than Hispanics. As I mentioned, confusion over how to do the blood kit was a reason given for a nonresponse and a reason to believe the in-person interviewing and specimen collection is likely to do better and finally the blood spot data do appear to have good validity.

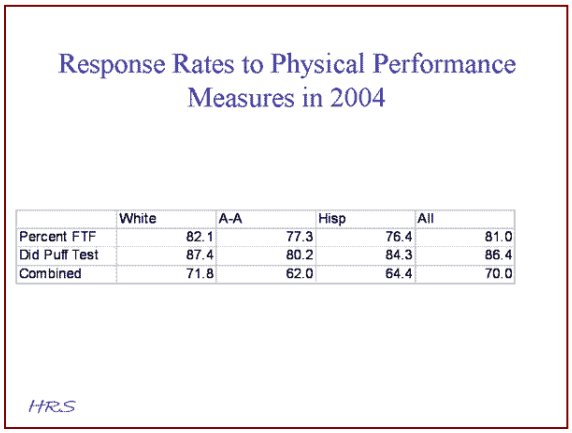


The last thing I want to talk about then is our physical performance measures that we conducted in 2004. We received some supplemental funding from primarily the Social Security Administration to do in-person interviewing to try and get higher consent rates to the linkage for social security. They weren't interested in physical performance measures, but we took advantage of the opportunity.

We drew a sample that was aimed at getting about 100 people to do these at every single year of age in the sample so it would be balanced in that way, and we'd have enough older people to do some things with.

Non-response is also a concern for these measures. There

were multiple ways to end up not doing the physical performance measures. One way is you could refuse to do the interview yourself, and we might get it via a proxy; and we wouldn't then collect any physical performance measure about you from the proxy. Another way is that you could say I don't want to do it face to face. I'll do it on the phone, and of course for the overall goals of the survey that's a lot better than nothing. Each of those was about 9 or 10 percent of the population selected to get physical measures, so that moved some people out and in some ways in a selected way which I'll show you.



Then once we're in the home and we're doing the in-person interview you have to sign a consent form, and about 6 percent of the people didn't want to do that; and then for every individual test you're asked, Do you feel comfortable with doing this test? And some people just ended up not doing a specific test.

Here are the overall response rates, using the puff test or the lung function and expiratory volume test as the primarily illustration. There was some minority differential in getting to the face to face. They were more likely to switch modes primarily, and so that got us some loss in the minority groups. And then given that we're in the home there was a differential again bigger for the African-Americans than for the Hispanics in willingness or whatever it takes to get you to do the lung function test once we're in the home and actually going through the interview there.

The combined rates are a differential of nearly 10 percent between whites and African-Americans and about 8 or 7 1/2 between whites and Hispanics. This loss of sample from all these various reasons is a continuing concern I think as we go forward. Now, remember again we have a 2 to 1 oversample, so we're still getting more than population representation even with slightly lower response rates from minorities, but nevertheless it would be nice to get them on as many as possible.

So one other dimension I want to look at as we look at this physical performance measures is functional status, and I'm not going to do this by race, but I think these are also an important considerations when you're talking about some of these biomarkers.

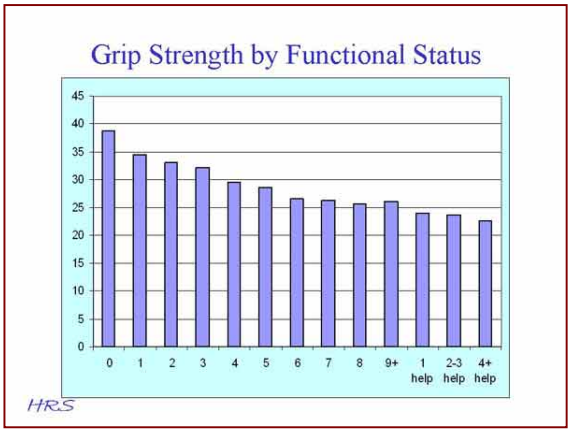
The horizontal here is a scale that I cobbled together from a combination of self-reported physical function limitations. That's things like can you climb up several flights of stairs? Do you have difficulty with steep kneeling, stooping or crouching? Do you have difficulty picking a dime up off the table? Do you have difficulty walking several blocks? There's about 12 of those that we ask every wave.

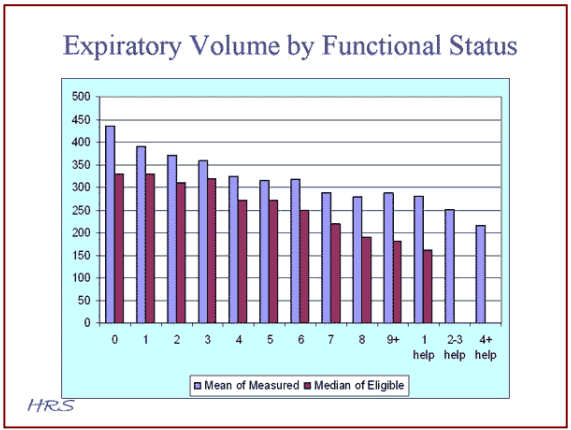
On this scale, a zero means somebody who says they can run a mile with no problem and no other problems; and somebody who's got three or four probably expressed some difficulty at kneeling or walking up several flights of stairs, so somebody who moves around, they're not in a wheelchair or anything, but they're beginning to have some limitation. Then the far end here is people who need help with ADLs, so the really extremely impaired population.

What you see is that through a combination of a higher rate of proxy interviews and a higher rate of refusal to do any individual test the response rates at the high end of disability are extremely low and I'm going to suggest probably very selective.

LINDAU: David, can I ask you a question about that? At what point do you think of them as a nonresponse versus I'd be willing to do it but I know I can't.

WEIR: That's absolutely the right question to ask because when you look at the performance what you see is a very steep decline in measured performance at those upper levels of self-reported function down to about the median, which is roughly three or four limitations. But from the median to the really low levels of self-reported ability, there's much less decline in this performance measure.





I think this is very much related to the fact that people who feel comfortable doing the test given their other disabilities are more likely to do the test than those who don't. For example, someone in a wheelchair with good hand strength might do the grip strength whereas someone with adequate lower body mobility but arthritis would not. This means we're not getting a representative sample of the people with severe physical impairments. So what I've done, and this is just very crude, is, instead of using the mean of those who did it, I assigned to the people who did not do it a value below the median and calculate the median, that's those red bars, what you see is something that looks much more like a continuously graded overall performance level. So that's a very crude way of doing it.

There's more sophisticated statistical approaches that would essentially take into account the probability that someone did this test as being correlated with their ability to do it and would make some corrections; but the point is I think one needs to think about doing that if one wants to relate performance on performance measures to functional ability when functional ability is a reason for not going the performance measures.

All right. So preliminary findings for the performance measures: response rates were generally good, but we may have given them too many opportunities to opt out.

There are still some minority differentials which gives us some concern going forward for how we're going to keep those response rates up. Response rates were lower for those with more levels of disability and so that selectivity of participation probably distorts the relationship between measured performance and self-reported limitations.

Okay. Thank you.

(Applause.)

LINDAU: We can take time for two or three questions if there are any. Sharon.

WILLIAMS: Did you lose any of the blood spot samples between the time they actually sent it in and analysis? Were there any that were sent in that were very poor quality?

WEIR: The protocol was for them to mail the sample directly to the lab -- we worked with a company called Flexite Diagnostics, who were very good to work with if any of you need to work with a lab that does that, they were terrific; and they had a mailer, so they did it. They put it on the paper, they put it in the envelope and mailed it themselves. So any loss between there and the lab we count as nonresponse so we just wouldn't know about that.

At the lab they then had a series of checks, and this is also what you learn as you do this work. They had a set of criteria for when they thought the sample would be valid which included the length of time between the time it was taken and the day it got to them to do the assay. We mailed in early October. Late in October we get a report back from some of the first results, and there's some fraction of these that they've ruled invalid because it had been too long; but their time interval was supposed to be like six weeks or something. Somebody had written down, I drew the blood in July. Well, they obviously hadn't because we only sent it in October.

So one thing you have to be sure you work with is what information to take as rock solid and what to question, so we asked them to go ahead and assay those that obviously had been done within the time window for validity, but ultimately there was some fraction of invalid, a couple percent I think.

DAVID: I'd imagine that people who were taking insulin would be much more comfortable giving a blood spot just because they were more familiar with pricking their finger several times a day. Did you find that to be true?

WEIR: We thought that, but it wasn't a strong pattern. It works both ways. On the one hand, yes, they're more familiar with doing it but on the other hand maybe they think oh, please, not another one. So it didn't seem to have a big overall effect.

HOUSE: I just wondered do you have any information on whether race of interviewer makes any difference?

WEIR: We don't yet. Certainly that's something that we want to pay very close attention to. We haven't done that yet for the 2004 data to see whether those physical performance response rates were higher in race matched interviewers than in unmatched interviewers.

WEINSTEIN: Bob Willis was concerned that asking for biomarkers would compromise future participation. Have you been back since the ADAMS?

WEIR: We have looked at the effects of supplemental studies on participation in the core. A colleague of ours, Bill Rodgers, has done most of this work. Everything we do that's new, we then follow up with an analysis of what happened to the response rates for the core survey the next time we went out; and so far nothing is deterring them in any differential way including the ADAMS. So far so good, and of course we'll find out in 2006 whether the face-to-face interview in 2004 in which we did the physical measures had any consequence for that.

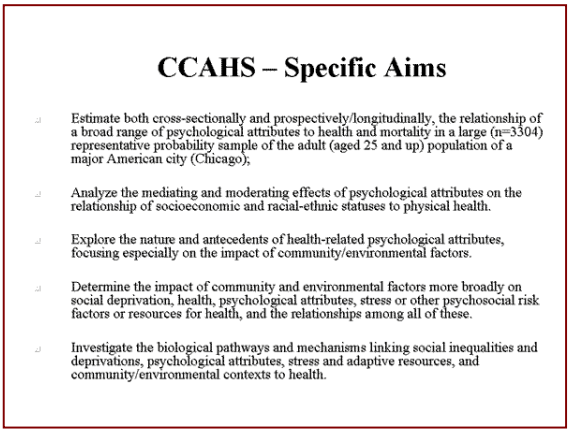
Again from what we heard the people who did it enjoyed it. It was a nice break. They said, Can we stop now and have you go away? They didn't want to go on and talk about their pensions after that, but doing the physical measures was fine. So it seems like it's good.

In fact six years ago we proposed to do some face-to-face interviewing only for the purpose of bonding people more closely to the study thinking that with years and years of telephone interviews they'd drift away, so we have reason to believe or hope that this will actually tie people into the survey, but we'll see as we push on it.

Chicago Community Adult Health Study (CCAHS): Biomedical Data Collection

James House

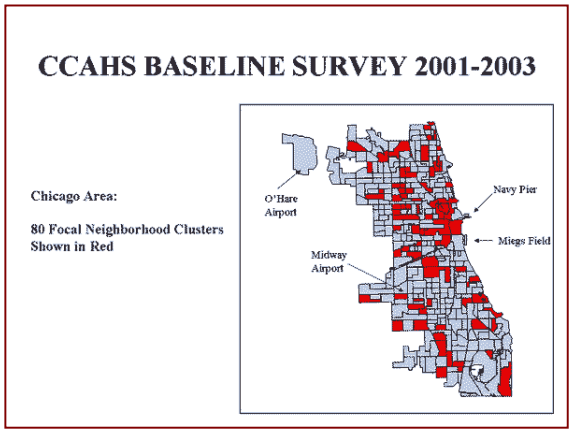
I will try to move quickly because I know people do want to go, and you can pick up other details which we'll try to fill in when we edit the other document. I'm going to talk today about our experience in something we call the Chicago community adult health study with doing various kinds of biomedical data collection, some within the in-home interview and some by follow-up attempts to collect blood and saliva. We have a lot of collaborators in this project so I'm talking on behalf of all of them.



This study was originally set up under a NIH mind body center grant. It had a set of foci that were on measuring psychological attributes in the community that were more specific to that. It has increasingly taken on as its major aims the goal of understanding the impact of neighborhood or community and environmental factors on people's living conditions, psychology, behavior, stress and ultimately on their health and as part of that we have tried to measure the health outcomes that we're interested in, both through self-reports and through doing some direct biomarker kind of collection.

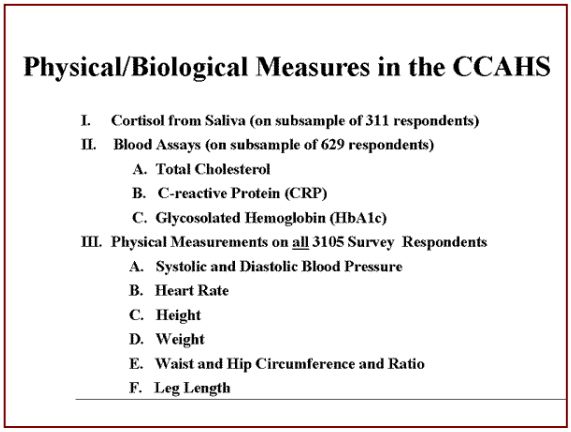
This just gives you an idea of the general framework of the studies and the kinds of things that are measured. This was the kind of research we were doing before we got into this looking at socioeconomic and racial ethnic disparities which are a major focus of this study and the way they may be explained by a whole set of psychosocial mediators and potential moderators that operate through biological pathways to produce longer disease outcomes.

The big addition or innovation of this study is to locate people within neighborhood context where we can measure both their sociodemographic characteristics of those context, what are sometimes called social advantage or disadvantage of those areas, as well as a whole range of specific factors or characteristics of those areas drawn from what we call the community survey where people characterize the areas that they live in and we aggregate the responses for people who live in a given area, something that we call systematic social observation that involves interviewers going around and rating and observing characteristics of the areas in which people live and then pulling together other information from administrative sources, and much of this data comes from census sources.



The study picks up off of the project on human development in Chicago neighborhoods that many of you know about, and what they did was to divide Chicago into 343 different so-called neighborhood clusters. Each of those is the size of about two census tracts. In addition for the purposes of their study they identified 80 of these areas, which are marked in red here, as so-called focal areas; and the focal areas were selected to try to achieve maximal socioeconomic variation within racial ethnic compositions. You tried to get predominantly African-American areas that had the full spread as much as possible across those areas in socioeconomic characteristics, the same thing for predominantly white areas and the same thing for a set of mixed areas.

They carried out a study of the development of children from adolescence to adulthood, longitudinal study, in these focal areas; and in those focal areas they also developed and -- innovated in developing something called systematic social observation which they did by having people drive through all of the streets in those areas and do some ratings on the spot while they were driving through and then videotape as they did that, and then they had raters go back and rate those things.



We essentially tried to adapt that process for more generic use in the surveys by developing a rating form that an interviewer can use to go around the block where a respondent lives and observe and mark the same kinds of characteristics.

So in the design of our study we were picking up studying adult health so we interviewed people 18 and over located in these areas. We oversampled people in the focal areas for two reasons: One was that those were the areas where the PHDCN and DCN study had the most information that we might be able to draw on, and then secondly these were the areas where we chose to pilot part of our biomedical or biomarker data and gave us some more concentrated areas that had good socioeconomic and racial ethnic distribution for trying that.

Okay. We did interviews that lasted about two hours with people. We did the systematic social observations, but I'll focus here on what we did in terms of what we call physical or biological data collection measures. Actually to take them in the opposite order they go on the slide, we did physical measurements, as we call them, in the actual interview on all of the respondents who would agree to do that. That included taking systolic and diastolic blood pressure, three measures using electronic blood pressure monitors, which also generate heart or pulse rate readings; and then we measured weight actually with scales, and we measured height, waist and hip circumference and leg length using a tape measure, standing people up and so forth.

To respond to Maxine's question yesterday, we tried to measure leg length directly by having people position the tape at the top of the hip bone and measuring down from there. We've just starting looking at that data. I thought it would be totally worthless, but actually it may not be. It does seem to be showing some reasonable relationships with other things.

As people have talked about, we had a range of successes and failures in doing this. At the end of the interview the people in the focal areas, the total sample was 3,105 which ended up dividing up into 1,145 people in the focal areas and 1,960 in the nonfocal areas. All of the 1,145 people in the focal areas at the conclusion of the interview were offered the opportunity to participate in a collection of saliva and blood.

The attempt in the saliva data collection was to leave behind with people eight Salivettes and to ask them to do four of those over two, quote, working days, consecutive working days in their week. In the case of the blood the attempt was to have the interviewer set up a time that a phlebotomist could return and collect a blood sample from them, and we had contracted with a medical personnel operation in Chicago for a nurse coordinator and a set of phlebotomists.

In doing this coming back, and we'll talk a little more about it, we were trying to be sensitive to the kinds of racial ethnic issues that have been discussed before. The interviewers covered a full range of African-American, Hispanic and white interviewers. Interviews were done in Spanish with Spanish-speaking people. Similarly we had at least one phlebotomist at all times who also was Spanish speaking and could be sent to Spanish-speaking households.

So this is what we were trying to do. We actually predicted or expected going in that we might get 60 percent cooperation of people who responded in the interview. The overall interview response rate for the total sample was 72 percent, which is pretty good in large urban areas at this point in time, so we were looking at 60 percent of that 72 percent.

So just to review quickly what happened for us, we made a design decision originally that we were going to avoid genetic issues for a lot of the reasons that were just talked about and because there had been a large flap, as you may recall, about genetics and violence during the middle 1990s out of NIH which occurred in the midst of the PHDCN study, and they advised us strongly to avoid that so we just said we weren't going to go there.

So in terms of the other measures, the physical measurements in the interview worked very well, better even than we would have thought. We had no major problems. In doing that you have some people who refuse body measurements, but we allowed people a number of options to be directly measured by the interviewer, to do the measurement themselves while being observed and instructed by the interviewer, and in extreme cases if they wanted to do it in privacy, we let them do it in privacy and report back the readings.

We used tapes that were measured in centimeters so that for most people it would not be quite so obvious what were measurements that were being taken when we're doing waist or hip, for example. We have some missing data that ranges from around 1 percent to 7.5 percent for the blood pressure readings; and the main problem we have after the fact identified and blood pressure readings is that those monitors are affected by the presence of arrhythmias, and that will give errors. There apparently are ways to get around them, but we didn't recognize the problem early enough to instruct people in that. So one could probably do even better and perhaps there may be improvements all the time in these monitors; but otherwise they worked extremely well, and that part of the study, as far as we can tell at this point, worked out very well.

LINDAU: Do you know if you used, did you have two cuff sizes like a medium and large?

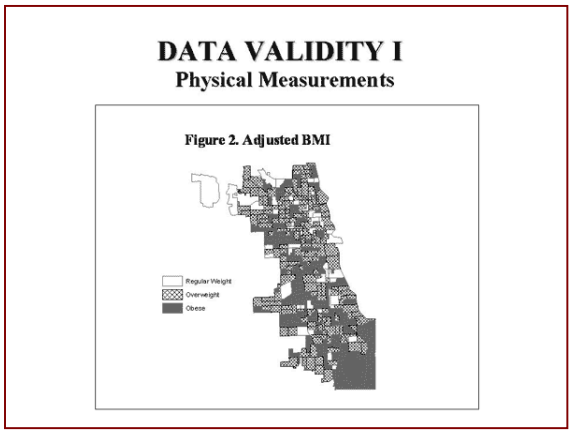
HOUSE: Yes, yes, we did that. I think there are like up to four different cuff sizes at the extremes. We had a regular and kind of a large, and so we had some problems. Again some of the measurements were lost with people who were extremely small or extremely, extremely large.

In the blood samples we ended up getting 55 percent successfully of people to do the blood samples. In general we had relatively little nonresponse differential by race and ethnicity in this. As best we can tell from our analyses thus far, people who cooperated were slightly older and had slightly more health problems so they may have seen these tests which we were going to report back to them as being more useful to them.

They were also slightly less likely to be or nonresponders were more likely to be Hispanic and first generation, and this may reflect again concerns that that population has either in dealing with these kinds of issues or in just repeated context where some of them may be worried about what the information may be going or what it's being used about; otherwise no major differences by race or socioeconomic status in cooperation. We tried to stress to people that this was being done both to give them information that would be useful to them and to provide information that ultimately might be useful to the community, and that seemed to work pretty well.

Cortisol, as someone said yesterday, it didn't break our heart, but it was extremely disappointing as an experience. We started out looking for four readings over two days. We encountered serious problems doing that. We then in our kind of refusal conversion follow-ups went back to trying to get four and then dropped ultimately to just trying to get two readings over two days, and in the end we basically have 311 people for whom we have a morning and an evening reading at least on one day.

The problems that occur in that were basically twofold. One is simple that you're actually asking -- we gave people instructions on how to do this, left the kits behind with them. It's a significant compliance problem for somebody to carry this out. We were not doing reminders. I think almost all the evidence that I've looked at since suggests you have to do reminders. That would have been -- you'll see the expenses and costs of this in a minute. That would have raised them to prohibitive levels for us.



The other problem we ran into in doing those is that people don't saturate the Salivettes sufficiently. We got back a lot of cases where people tried to comply, but there was not enough saliva in there to do it. I'll show you it may not be a total loss, but I thought going in this should be fairly straightforward. It turns out it's not to do that.

Just to look a little bit at some evidence that the measures work reasonably well, we've been doing some analyses of the height and weight measures and have cleaned them up, and they work pretty well. This is an image of the mean weights in all of the areas as to whether they fall into the obese, overweight or regular weight category in Chicago. Chicago overall is a heavy city. The overall BMI is about 29.5 as I recall. You can see how many areas there are that have means that are in the obese range, and you can see that those -- that there is a tendency of those to cluster in areas that are either highly populated by the, quote, racial minority populations or lower socioeconomic populations or both.

These are the kinds of disparities that one observes. There are differences that range. These are for females where they're greatest. They're similar for males but smaller. We're talking 3 to 4 point differences between Hispanics and -- or 3 to 4 1/2 point differences between Hispanics and non-Hispanic blacks versus whites in this population.

Similarly by education we're looking at differences that are 2 1/2 to 3 points, and these are substantial differences on BMI. Income tends to be something that differentiates primarily within the white population, not as much in the African-American Hispanic population so overall it has much smaller affects.

Blood Pressure

Hypertensive Means (and Standard Errors) by Sex and Race/Ethnicity in the CCAHS				
	Males		Females	
Black	37%	(2.3%)	28%	(1.5%)
Hispanic	31%	(2.2%)	19%	(1.8%)
Other	23%	(5.5%)	24%	(4.9%)
White	25%	(2.2%)	18%	(1.8%)

This analysis, I'll just summarize, was an attempt to get an estimate of what proportion of the total differentiation that we see by race and ethnicity or socioeconomic status may be accounted for by other characteristics like birth history, immigrant generation at the individual level; and you could compare these numbers and see there's probably a 25 to 30 percent reduction in the size of the coefficients without adjustment and then to see of this remaining individual level of variation, how much of that might be accounted for by contextual variation, that is, how much of it lies between the neighborhood clusters rather than within. We do that by a fixed effects model, and you can see in general about 30 percent, plus or minus, of the variation here looks like it may be a function of the context that people live in rather than characteristics of them as individuals or households. We're now working to try to figure out what those are.

Blood pressure variation in the population also looks pretty reasonable. As you might expect, it's highest among African-Americans and Hispanic males and lower in the white population and the small group of other race ethnicity groups. This is a very busy table, but it does show if you look closely at it that there are substantial racial ethnic differences for HbA1c measured from the bloods collected and in C reactive protein. There are not so clear racial ethnic differences, and the same is true socioeconomically you can see that are some fairly greater differences. HbA1c is greater than 6 percent; go down pretty much linearly with education and so forth.

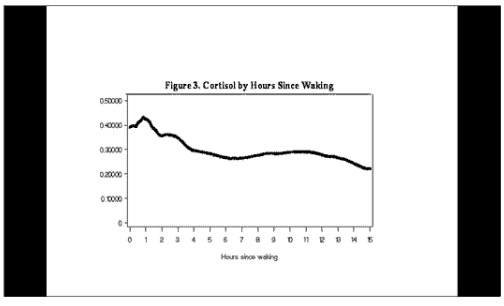
DATA VALIDITY II
Blood Draw Assays

HBA1c, CRP, and Total Cholesterol by Race/Ethnicity and Education Among Females and Males

	Race/Ethnicity								Education					
Females	Non-Hispanic Blacks		Non-Hispanic Whites		Non-Hispanic Other		Hispanic		0-12		12-15		16+	
HbA1c	5.9	(0.10)	5.2	(0.11)	5.3	(0.27)	5.6	(0.11)	6.0	(0.12)	5.5	(0.08)	5.3	(0.12)
HbA1c>6%	4.4%	(0.6%)	1.2%	(0.8%)	1.7%	(2.1%)	3.2%	(0.7%)	5.5%	(0.8%)	2.7%	(0.6%)	1.1%	(0.8%)
CRP	0.69	(0.06)	0.38	(0.07)	0.33	(0.17)	0.60	(0.07)	0.55	(0.08)	0.63	(0.05)	0.38	(0.08)
CRP>1.0	19.6%	3.0%	7.0%	3.2%	0.5%	8.2%	14.5%	(0.03)	13.9%	3.7%	15.7%	2.5%	7.8%	(3.8%)
Chol	192	(3.3)	207	(3.6)	185	(9.1)	193	(3.8)	194	(4.1)	198	(2.8)	196	(4.3)
Chol>200	7.5%	(1.0%)	11.5%	(1.2%)	7.9%	(3.4%)	8.0%	(1.2%)	9.0%	(1.3%)	9.4%	(0.9%)	7.1%	(1.4%)
Males	Non-Hispanic Blacks		Non-Hispanic Whites		Non-Hispanic Other		Hispanic		0-12		12-15		16+	
HbA1c	5.9	(0.10)	5.2	(0.11)	5.3	(0.27)	5.6	(0.11)	6.0	(0.12)	5.5	(0.08)	5.3	(0.12)
HbA1c>6%	6.1%	(1.0%)	0.9%	(0.9%)	5.1%	(2.3%)	3.6%	(0.5%)	3.7%	(1.1%)	3.4%	(0.7%)	3.8%	(1.0%)
CRP	0.69	(0.06)	0.38	(0.07)	0.33	(0.17)	0.60	(0.07)	0.55	(0.08)	0.63	(0.05)	0.38	(0.08)
CRP>1.0	11.8%	(2.7%)	1.9%	(2.4%)	0.4%	(5.3%)	4.3%	(2.4%)	7.7%	(2.9%)	5.8%	(2.0%)	2.2%	(2.6%)
Chol	192	(3.3)	207	(3.6)	185	(9.1)	193	(3.8)	194	(4.1)	198	(2.8)	196	(4.3)
Chol>200	8.5%	(1.5%)	8.2%	(1.4%)	13.1%	(3.6%)	9.7%	1.4%	8.1%	(1.8%)	7.7%	(1.1%)	10.2%	(1.6%)

All in all the blood data and the blood pressure and physical measurement data to the point that we've analyzed them appear to be working in the way that you would expect suggesting they have reasonable construct validity.

DATA VALIDITY III
Cortisol



The cortisol data, about all we can say at this point is we have looked at the diurnal rhythm, and in spite of the problems we've had and so forth, we get a diurnal rhythm that is not too far off of what one expects normally. The hope is that we may yet be able to derive something from this so that we would want to do it again in a different way.

I thought I'd look at some cost numbers because for at least those of us who do large scale surveys, that's a big issue in terms of the addition of biomarkers. To do the survey part of this, which was about a two-hour operation, the cost of that was approximately $735 I guess it was per interview. It was about 2.25 million total, which is a lot of money these days.

Of that the physical measurements took about one-eighth or seventh to eighth of the interview; cost about $90. My assessment would be that that's actually quite cost effective; that being able to get that kind of information directly given the marginal cost of once you've gone to the trouble of getting somebody there to do an interview for other purposes to collect those kinds of data was relatively inexpensive.

Bottom Line: Complicated, Costly, and Variably Cost –Effective

- 1. Approximate cost of survey interview (excluding physical measurements) = $642/Respondent
- 2. Approximate cost of Physical Measurement portion of interview = $93/Respondent
- 3. Approximate cost of Blood & Saliva portion of interview = $239/Respondent (includes indirect costs, so direct costs ≈ $480/Respondent)

There's a big error here. I tried to make a last-minute adjustment. This number, the total cost here should be the sum of these two numbers which is $719, and it cost essentially as much in total cost to us to do the blood and saliva data collection on a person as it did to do the survey and physical measurement.

Now, that is the main part of that cost, and what I was trying to do here was to divide it because we were paying a contractor and do direct cost so that the fairer comparison maybe $500 per person to do this direct cost plus an indirect cost payment of 239 if that was a rough estimate of what their total charge might have been if they were using indirect cost structure similar to ours.

So it cost somewhere between 50 and 75 percent of the cost of the interview and physical measurement to do that. The bulk of that was eaten up by the process of trying to get a phlebotomist hooked up with a person; that people simply don't show up. They say they're going to do it. They don't respond. You have to keep calling them and going after them. Once you make that connection, the rest of the cost is not that bad. Phlebotomists took the blood, put it on ice, took it to a lab, tests get run and so forth; but the bulk of that cost is in that process.

So clearly the kinds of things that are being talked about here to try to do this process directly during the interview and to save, and I was disappointed in some ways to hear that you guys are going to have to go through the process of trying to get phlebotomists out there because at least in our experience in a major urban area that is a big hassle and a big cost.

Looking to the Future

- BP and Physical Measurements Relatively Straightforward and Cost Effective (cost-effectiveness of weighing and measuring relative to self report worthy of more study)
- Blood collection/assays could be improved and made less costly by less invasive technique, requiring less special personnel and handling. Other labs/providers could improve even the process we used. Validity of former *vs.* latter an issue
- Cortisol uncertain without follow-up reminder process (which may or may not be cost effective)
- May explore genetics with a small subsample of already cooperative Rs

So kind of in conclusion and looking forward our sense was that certainly the experience of doing the blood pressure and the physical measurements was really one that was fairly straightforward and quite cost effective. I think one might do some study. I think David mentioned he was doing a little bit of investigation of this.

My epidemiologist colleagues tell me, yes, you have to carry scales out there and weigh people because they lie to you. I'm not sure the gains. Again that's the most costly and cumbersome part of that besides buying the blood pressure monitor itself of that process. It's not quite clear to me that since the errors tend to be out at the extreme end of the distribution where it's not going to affect your relationships with other things that much that we may need to do that, but that's kind of at the margin a small improvement.The blood collection and assay part obviously could be vastly improved by use of these less invasive techniques if they can work successfully and using fewer personnel and involving less handling of stuff. If we used other providers, we could have improved greatly not even but over the process that we used; and I think Tom mentioned yesterday that the L.A.FANS does have to go through this

process. There might be some useful things that could be gained by collecting serum and spots on some people and comparing for those who are concerned about the validity of blood spot methods that would be a real opportunity to do that in a general population.

Cortisol at least in my experience I wouldn't try to do in a large population survey again unless I could implement a reminder process and did a better training process to begin with, and that raises serious cost issues. That still could be done in subsamples. It would be lovely if we could figure out a way to make use in a better way of a single measure that could be collected while a person is there, but I think we know there are all kinds of problems.

Genetics we shied away from I think from what I heard probably appropriately at the beginning because we didn't want to jeopardize the rest of the operation by getting into that. We are thinking at this point about going back to people who have already been cooperative in biomedical samples and trying for that subgroup of people saying would you be willing to participate in a study that involved collection of genetic material.

So that's the general picture of our experience, and I'd be glad to answer any questions.

(Applause.)

LEITSCH: You mentioned very early in your talk that you gave the respondents the option to perform their own physical measures, and I was wondering if you had taken a look at how frequently the respondents took you up on that and whether it affected data quality or not.

HOUSE: We have an indicator. I'm pretty sure that is -- we have an indicator. In a few cases they self-reported. In some cases they self-reported information to us and in some cases they didn't, and I believe the interviewer indicated that. I just can't tell you the data, and we haven't looked at it yet, but it's a good question.

SASTRY: Can you say just a little bit more about how the hand-off from the field interviewers to the phlebotomists occurred in terms of who did the initial scheduling and where the trouble sort of started?

HOUSE: What we tried to do was to get the interviewer to make an appointment, to get an appointment from a respondent for a period about two to three days down the road to give enough time for the hand off to occur. The problems just begin at that point. Some people won't give you an appointment so then they get passed off to the nurse coordinator to call them back and do it.

We'd have to go back through all of our call record data to figure out how many, but there were a lot. All I can remember is there were a lot of cases where the appointment was broken in some way initially and whereupon the coordinator would have to try to recontact these people, set up another appointment, and get the phlebotomist out there and so forth. It's that whole process.

It's very much the same as interviewing where again the bulk of the cost is actually generated by getting the interviewer into the home with the respondent, and then the marginal cost at that point is not at all that bad.

ADAM: Of the 311 that completed the core, how many did you ask to complete it?

HOUSE: Depending how you want to start. We started there were 1,145 people. Of those about 600 and some, 686 I think it was said that they would do biomedical samples and were left with the kits. Of those probably, as I recall, around 400, it's in the upper 400s actually did return, mail back to us Salivettes; and of that after we did the lab analysis, we end up with this kind of 311 minimal usable cases.

The only good news in all of that is that as best we can tell, that seems to be a fairly random process so that we're not seeing systematically that the data work came differentially by race, ethnicity or socioeconomic.

ADAM: If you do try it again, in terms of (inaudible) schedules, what I found to be most effective is a call the evening before the morning they're supposed to do it; and then if you could afford it, a follow-up call the evening they were supposed to complete it to get them to put it in the mail the next morning; and we get over 90 percent with that.

HOUSE: I've read other papers now that suggest that that's the case; and we actually thought of doing that, but our field people, they were struggling in all kinds of ways pulling the study together. It seemed to be over the top to ask them to do one more thing like that.

ADAM: The other effective thing is to actually have them do the first sample with the interview even as a throw away sample just so they've gone through the procedure. That way they don't have to read the instructions.

HOUSE: Absolutely, I would agree with both of those in retrospect are totally right on.

LINDAU: Are you enthusiastic about doing biological data collection again in a future study?

HOUSE: Yes. In terms of when you actually get the data and start to use it, as far as we can tell at this point in time, I'd say with the cortisol kind of pending just because it got reduced so far that we don't have a lot of power left for analyses, yes, it's there; and I think it's proving already in very early stages to be useful.

On the other hand I think we've got to find some ways to do this in large scale population studies. We've got to find and demonstrate that methods that are somewhat more cost effective can work. Blood pressure is a case where again the advance and the technology for the measurement there has now made that a kind of a nonproblematic thing; whereas previously you would have had to have a much bigger training process or experience personnel sent out to do it and so forth, and that's been solved. I'm hoping that what Thom and other people are doing will solve the blood pressure problem as well.

It hasn't in any way diminished our enthusiasm. I just put it out as a note for consideration, and particularly again I think the more that we can develop ways of doing this successfully within the same contact with the time with the person, the much greater success and ability we will have to do this on a large scale.

Closing

Thomas McDade

I'm going to be very brief for those that are still left standing, and I'm barely among you. I want to start by thanking April and Maria and Nathan for logistical help over the past couple days, especially with the microphone duty. I'd also like to thank NIA for their support of the workshop, our colleagues from NIA for coming out and spending a couple days with us and participating in the workshop, and NIA more generally for taking a leadership role in helping us implement biomarkers into a lot of the studies that we've been talking about over the past couple of days.

We've covered a lot of territory. Literally, we started with Taiwan. We've been to the Philippines. We've been to Los Angeles. We've been to Germany, and we've even been to outer space. That's pretty exciting and I think marks some interesting new directions.

A lot of early biomarker research was started overseas, motivated largely by the conditions that a lot of us in public health and anthropology have faced in terms of collecting samples. So it was great to see some of that represented here more centrally in this workshop.

We've also talked a lot about aging, but we've also talked about adulthood, young adulthood, adolescence, childhood and even infancy and prenatal environments, which is also very exciting and reflects the reality of health as a process that is a product of the life course and life cycles. I was excited about that direction in this year's workshop, and I look forward to seeing that continue as well.

Lastly, I'd like to end with where we started, which was John Cacioppo's discussion of the process of science and how we can do innovative, ground- breaking collaborative research, and some of the pitfalls associated with that. I've thought a lot about his presentation since then, which was not the talk that I expected him to give; but I think it was a wonderful way to start, and I've been thinking a lot about how we can try to work together collaboratively to generate some dynamic synergy that can lead to some interesting breakthroughs. We're poised on the precipice of some very exciting opportunities here, and I see this workshop as a great way to facilitate that process. Hopefully you do as well.

So on behalf of Stacy and her colleagues at the University of Chicago, on behalf of my colleagues at Northwestern, I thank you for coming and for participating. This really is a wonderful collective effort, and I look forward to this meeting every year, and I look forward to seeing you all next year. Thank you for coming and save travels home.