CCBAR

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

The 2004 Chicago Workshop on Biomarker Collection in Population-Based Household Surveys of Older Adults

Proceedings

Acknowledgements

The 2004 Chicago Workshop on Biomarker Collection in Population-Based Household Surveys of Older Adults was organized and coordinated by Dr. Stacy Tessler Lindau and Dr. Jenna Mahay, with assistance from Kathleen Parks, Jessica Lester, Adelle Hinojosa, and many others.

Financial support for the 2004 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.

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 0-9769978-1-9

Suggested Citation

Lindau S, Mahay J (Eds). Proceedings of the 2004 Chicago Workshop on Biomarkers in Population-Based Health and Aging Resarch. Center on Demography and Economics of Aging, NORC at the University of Chicago. Chicago, IL. May, 2005.

Preface

The traditional clinical medical research model promotes discovery primarily in the realm of illness rather than in the maintenance of health or prevention of disease. Increasingly, biomedical scientists conducting human-level health and disease research and social scientists studying health and illness are meeting and collaborating in the population laboratory.

Major advances in survey research technology and infrastructure, in combination with the development of methods to enable collection of biological and physiological data in the home setting, present a new opportunity for studying health and illness in vivo, involving probability-based, representative samples of a population. Integration of social scientific with biomedical methods advances health research beyond critical sampling biases introduced by clinic-based research and advances population-based health research beyond subjective or self-reported measures of health. Integrated, population-based health research offers a most powerful tool to address health disparities and to reach remote populations by including individuals who cannot or do not access clinics and hospitals for health and medical care.

In 2004, the National Institutes on Aging instituted 5 years of support for the development of a "think-tank" at the Center on Demography and Economics of Aging Core on Biomarkers in Population-Based Health and Aging Research (CCBAR), NORC at the University of Chicago. The activities of the core are three-fold: 1) to facilitate interdisciplinary discourse and collaboration among biomedical and social scientists working in the population laboratory to study health and aging, including establishment of a new website (http://biomarkers.uchicago.edu); 2) to train social science post-doctoral, research, and junior faculty in basic principles of human biology and physiology; 3) to host an annual workshop on issues pertinent to integrated population-based health and aging research.

This publication summarizes the proceedings of the Second Annual Chicago Workshop on Biomarkers in Population-Based Health and Aging Research. It contains edited transcripts and excerpts from slide presentations given at the 2004 workshop held June 10-11 at the University of Chicago Gleacher Center. The purpose of this document is to provide a resource and reference for individuals contemplating or engaging in the collection of biological and physiological data in combination with social survey data in the population setting. Financial support for the 2004 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.

The 2004 presentations covered ground in three main territories: 1) theoretical foundations and core challenges, including human subjects and ethical considerations, of collecting integrated social and biological data in population-based research; 2) innovations in measurement, particularly minimally invasive measurement, of biological and physiological factors; and 3) a continuation from 2003 of the discussion on studying cognitive function and impairment in the population laboratory (see *Workshop on Biomarkers in Population-Based Health and Aging Research*. 2003. Chicago, IL). The lunchtime "Translations" session featured a speaker on forensic science from the FBI, Special Agent Douglas Hyde. He 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.

To accomplish integrated health and illness research in the population setting, many areas of need exist. These include development of methodology for analytic integration of biological and social data as well as methods for streamlining collection of data in the population setting. The latter requires: 1) advances in instruments to obtain reliable and valid self-report data; 2) development of minimally invasive techniques for collecting biological and physiological data; 3) establishment of best practices for integrated health research including training of lay interviewers to collect such data and clarification of ethical and human subjects issues, and; 4) establishment of a national laboratory or network of laboratories capable of interacting securely and efficiently with population-based health researchers. In addition to advancing human-level health research, CCBAR aims to identify and implement strategies for translation of research methods to improve scientific inquiry as well as medical diagnosis and treatment in remote or understudied areas and in populations with limited access and/or mobility.

We hope you find these Proceedings informative. They can also be found online at:
http://biomarkers.uchicago.edu

Stacy Lindau, MD, MAPP
University of Chicago Pritzker School of Medicine and
Center on Demography and Economics of Aging, NORC at the University of Chicago
May 18, 2005

List of Participants

Brauner, Daniel	Associate Professor, Geriatrics Section, Dept. of Medicine Fellowship Director, Geriatrics University of Chicago
Chatterji, Somnath	WHO Multi-Country Studies, Study on Global Ageing Measurement and Health Information Systems
Clark, Christopher	Director, Memory Disorders Clinic Associate Director, Alzheimer's Disease Center University of Pennsylvania
Coelen, Craig	President, NORC at the University of Chicago
Friedman, David	Associate Professor of Ophthalmology and International Health Wilmer Eye Institute and Johns Hopkins Bloomberg School of Public Health
Garfield, Craig	Instructor of Pediatrics, Children's Memorial Hospital Evanston Northwestern Healthcare Research Institute
Gavrilov, Leonid	Researcher, Center on Demography and Economics of Aging NORC/University of Chicago
Gavrilova, Natalia	Researcher, Center on Aging Researcher NORC/University of Chicago
Hahn, Anneliese	National Institute on Aging Behavioral and Social Research
Harris, Kathleen Mullan	Professor, Department of Sociology Director, National Longitudinal Study of Adolescent Health Carolina Population Center University of North Carolina at Chapel Hill
Herndon, Mitch	Program Manager American Social Health Association (ASHA)
Hughes, Mary Elizabeth	Assistant Professor, Department of Sociology Duke University
Hyde, Douglas	Special Agent, FBI
Jaszczak, Angie	Survey Director NORC at the University of Chicago
Koppelman, Missy	Vice-President/Associate Director Health Studies, Biomarker Labs NORC at the University of Chicago

Kuby, Alma	Senior Survey Director II, Biomarker Labs NORC at the University of Chicago
Kurina, Lianne	NIMH Postdoctoral Fellow Department of Health Studies University of Chicago
Langa, Ken	Assistant Professor of Internal Medicine Assistant Research Scientist, Institute of Gerontology University of Michigan/and Institute for Social Research Department of Internal Medicine University of Michigan
Lantos, John	Associate Professor of Pediatrics, University of Chicago Associate Director, MacLean Center for Clinical Medical Ethics University of Chicago
Latta, Federica	Department of Medicine University of Chicago
Laumann, Anne	Assistant Professor, Dermatology Section, Department of Medicine University of Chicago
Laumann, Edward	Professor, Department of Sociology University of Chicago
Leitsch, Sara	Research Associate (Assistant Professor), Department of Obstetrics and Gynecology University of Chicago
Lindau, Stacy	Assistant Professor Depts. of Ob/Gyn and Medicine (Geriatrics) University of Chicago
Lipton, Rebecca	Associate Professor, Endocrinology, Pediatrics University of Chicago
Lorincz, Attila	Senior Vice President, Research and Development Digene Corporation
Lundeen, Katie	Survey Specialist NORC at the University of Chicago
Lundstrom, Johan	Ph.D. Candidate, University of Chicago McClintock Lab, Institute for Mind and Biology
Mahay, Jenna	NIA Postdoctoral Research Fellow Center on Demography and Economics of Aging
Masi, Christopher	Assistant Professor of Medicine Section of General Internal Medicine The University of Chicago

McClintock, Martha	Director, Institute for Mind and Biology Professor, Department of Psychology University of Chicago
McDade, Thomas	Assistant Professor, Department of Anthropology Director, Laboratory for Human Biology Research Northwestern University
McPhatter, Tarnya	Program Manager American Social Health Association (ASHA)
McQuillan, Geraldine	Center for Disease Control and Prevention National Center for Health Statistics
Moore, Christine	Laboratory Director US Drug Testing Laboratories
O'Muircheartaigh, Colm	Professor, Irving B. Harris Graduate School of Public Policy Studies, The University of Chicago Vice-President for Statistics and Methodology at NORC
Parish, William	Professor and Chair, Department of Sociology
Parks, Kathleen	Administrative Director, Academic Research Centers NORC at the University of Chicago
Patterson, Elissa	Institute for Mind and Biology Department of Psychology University of Chicago McClintock lab
Pham-Kanter, Genny	Doctoral Student, Departments of Sociology and Economics University of Chicago
Pierret, Charles	Director, National Longitudinal Surveys Bureau of Labor Statistics
Plassman, Brenda	Associate Research Professor Director, Program in Epidemiology of Dementia Department of Psychiatry Duke University Medical Center
Pollack, Harold	Associate Professor, Social Service Administration University of Chicago
Raina, Parminder	Director, McMaster Evidence-based Practice Center Associate Professor, Department of Clinical Epidemiology and Biostatistics, McMaster University
Rhodes, Karin	Director, Health Services Research Group Section of Emergency Medicine University of Chicago

Sachs, Greg	Associate Professor of Medicine Chief, Section of Geriatrics University of Chicago
Savage, Cathleen	Vice-President, Human Resources and Administration NORC at the University of Chicago
Schumm, Phil	Statistician, Department of Health Studies The University of Chicago
Seeman, Teresa	Professor of Medicine and Epidemiology, Division of Geriatrics Geffen School of Medicine at UCLA
Smith, Stephen	Senior Survey Director NORC at the University of Chicago
Waite, Linda J.	Lucy Flower Professor, Department of Sociology Director of the Center on Aging, University of Chicago
Wallace, Robert	Professor of Epidemiology and Internal Medicine University of Iowa
Williams, Sharon	Visiting Professor, Department of Anthropology Northwestern University
Willis, Bob	Professor, Department of Economics Research Professor, Survey Research Center, University of Michigan Survey Research Center
Wolfson, Christina	Professor, Department of Medicine, Department of Epidemiology and Biostatistics, McGill University Director, Centre for Clinical Epidemiology and Community Studies, Jewish General Hospital
York, Erin	Project Coordinator, National Social Life, Health, and Aging Project (NSHAP) Doctoral Student, Department of Sociology University of Chicago

Chicago Workshop on Biomarker Collection in Population-Based Health Research

At the Gleacher Center, University of Chicago
Agenda June 10th and 11th, 2004

Thursday, June 10th: The New Frontiers of Biomarker Collection Technology

8:30 – 9:00 am	**Breakfast reception**
9:00 – 9:15 am	**Introduction and Welcome to the Workshop**
	Stacy Lindau
9:15 - 9:30 am	**Overview of Interdisciplinary Work to Uncover the Mechanisms Linking Social Life and Health: Core Challenges**
	Teresa Seeman
9:30 - 9:45am	**Introductions by Participants**
9:45 - 11:45 am	**New Developments in Minimally Invasive Biomarker Collection Technology** The purpose of this discussion is to 1) share information on new developments in biomarker collection technology; 2) discuss the continued challenges for collecting certain types of biomarkers in population-based research (particularly with non-medically trained interviewers); and 3) suggest future directions for the advancement of biomarker collection technology (what new technologies need to be developed?).
	Moderator: Geraldine McQuillan Jenna Mahay Thom McDade Christopher Masi Christine Moore Attila Lorincz Martha McClintock
12:00 - 1:15 pm	**Lunch & Keynote Speaker Douglas Hyde: "Translations from Forensic Science: Perspectives from an FBI Special Agent"**

1:30 - 3:00 pm	**Studying Cognitive Function in the Population Setting: Possibilities and Limitations** The goal of this discussion group is to bring together diverse perspectives on how we could best measure cognitive function in population-based research and what else we would need to measure in order to decipher the mechanisms linking cognitive function to physical, mental and socioeconomic health and well-being.
	Moderator: Ken Langa Chris Clark Brenda Plassman Bob Willis Robert Wallace Dan Brauner Discussant: Greg Sachs
3:00 - 4:30 pm	**New Developments in Measuring Sensory Function in Population-Based Research (and what can they tell us about health?)** This discussion group will present newly developed methods for testing sensory function in population-based studies, and discuss what they can tell us about cognitive function and physical health. Is sensory function important for understanding the link between social life and health?
	Moderator: Sara Leitsch Johan Lundstrom David Friedman Erin York Sharon Williams Federica Latta

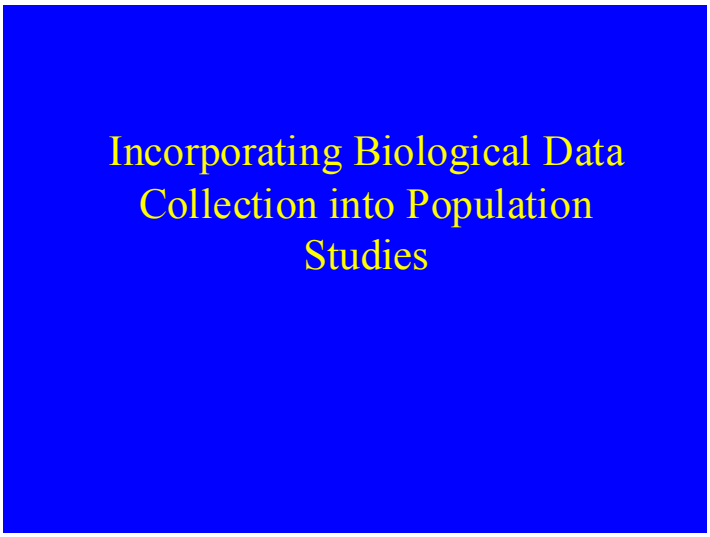
Friday, June 11th: The Human Side of Biomarker Collection

8:00 – 8:30 am	Breakfast
8:30 – 10:00 am	**Biomarker Collection in Population-Based Health Research: Human Subjects Issues** What are the most pressing human subjects issues faced by researchers collecting biomarkers in population-based studies? This group will share their diverse experiences with issues such as IRB applications, informed consent, risk, liability, insurance, and results notification and counseling protocols.
	Moderator: Alma Kuby Kathleen Mullan Harris Parminder Raina Cathleen Savage Rebecca Lipton Tarnya McPhatter Karin Rhodes
	Coffee Break

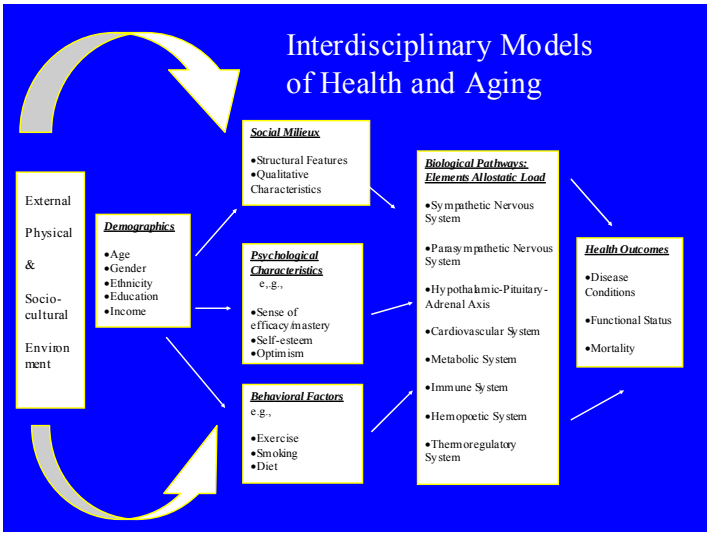
10:15 - 11:30 am	**Honest Brokers:** What are they, and do we need them? This discussion will focus on the issue of honest brokers: What are they, how do they work, do we need them, and do they really protect human subjects?
	Moderator: John Lantos Phil Schumm Stephen Smith Kathleen Mullan Harris
11:30 am - 12:00 pm	**The Big Picture: What is biomarker collection good for?**
	John Lantos: 1) Does the honest broker system really improve human subject protection? 2) How to keep sight of the ultimate goal of improving health through biomarker collection in population-based research? 3) How do we anticipate and avoid pitfalls that could result in this ground-breaking work not resulting in improved health?

Incorporating Biological Data Collection into Population Studies

Speaker: Teresa Seeman



I was asked to give an introduction and a little background on how many of us got to the point of being interested in this question of incorporating biomarkers into some of the more population-based studies.



This is my own illustration of some of the complex variables and pathways that we've gotten interested in. But the main thing this is meant to illustrate, at least for the purposes of today, is that part of what's happened is a lot of the traditional large population studies which had historically looked at more demographic and behavioral and psycho-social kind of factors in relation to health outcomes like functioning and mortality, have now gotten interested in trying to understand the biological pathways. I think that's a major underlying motivation for many of us in trying to get assessments of some of these biological parameters, is to try to nail down a little better some of the biological pathways through which these variables, the psycho-social and the behavioral, operate in terms of their impacts on disease and functional outcomes that we've studied more traditionally.

Candidate Biological Parameters

- Cardiovascular Factors
  - Systolic Blood Pressure
  - Diastolic Blood Pressure
  - Pulse
  - Heart rate variability
  - Response to challenge
- Metabolic Factors
  - HDL Cholesterol
  - LDL Cholesterol
  - Total Cholesterol
  - Glycated Hemoglobin
  - BMI/WHR
  - Fasting Triglycerides
  - Fasting Glucose (and/or OGTT)
  - Fasting Insulin
- Sympathetic Nervous System
  - Epinephrine
  - Norepinephrine
- Hypothalamic Pituitary Axis
  - Urinary/Salivary Cortisol
  - DHEA-S
- Inflammation
  - Serum Albumin
  - CRP
  - Fibrinogen
  - Il-6
- Genetic Differences

I've tried to list here some of the candidate biological parameters that have been of current interest in many areas. Cardiovascular factors are more traditional, blood pressure, pulse. Heart rate variability is a relatively new protocol that we're trying to move into more population-based kind of work. A range of metabolic parameters are frequently examined. The sympathetic nervous system and hypothalamic pituitary axis function have been assessed in terms of urinary and salivary cortisol assessments, and, most recently, markers of inflammation have gained more prominence. This is not meant to be comprehensive, but just to show you, in large measure, the range of systems that have been of interest and the wide range of different parameters that have been incorporated into various studies.

Just to give you a little background on my own perspective on how this has evolved, I think the domain where you've seen the longest history of incorporating actual biological assessments into population-based kinds of research as opposed to just clinical work is in epidemiology. This is my background, which may be why I have the bias of viewing that as having historically been a player in this area. There it's largely been in studies like the Framingham study which many of you may be familiar with, the more recent Cardiovascular Health Study, which is also population-based, and of course, Michael Marmot's Whitehall study. The main thing that I want to point out with respect to most of these studies is that the one big difference from what we're all interested in today is that those studies have largely been based on population samples, but they brought people into clinics to do the assessments, which means they didn't have to deal with a number of the logistical and feasibility kinds of issues we will. So that's the background, but they did do this work based in clinic settings.

Historical Role of Biological Assessments

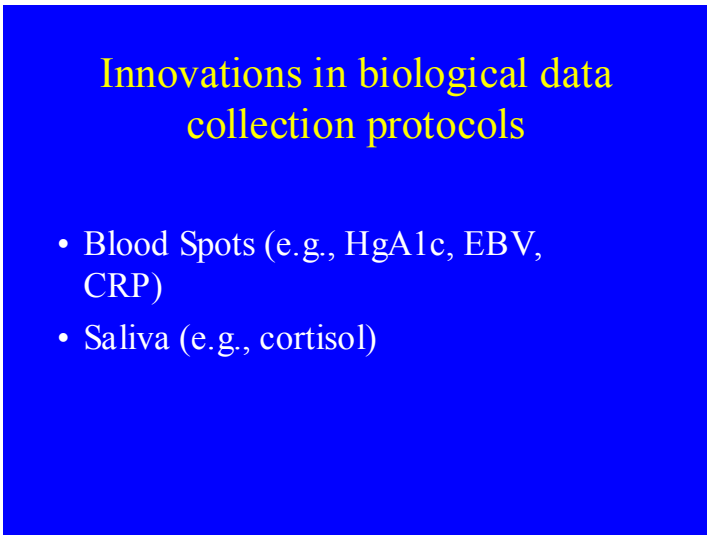
- Epidemiological studies with population-based sampling but clinic-based data collection
  - Framingham
  - Cardiovascular Health Study (CHS)
  - Whitehall

Development on more Field-based Approaches

- MacArthur Aging/EPESE
  - Blood
  - Urine
- Nurses Health Study
- "Mobile Clinics" (NHANES, HANDLS)

More recently, there have been a number of attempts to move some of this biological sampling out into the community. Again, I make the disclaimer here: this is not meant to be comprehensive. If others know of other studies that I haven't mentioned, no one should take that as ignoring them. I'm most familiar with the MacArthur Aging Study and the EPESE where we actually went out and did blood samples and urine samples in the community. The other one I wanted to mention is the Nurses' Health Study. I don't

know how many of you are very familiar with it. It's a prospective longitudinal survey of nurses across the U.S., and what they've done -- which I thought was really quite clever but may not be feasible in all cases – is one of the ways they collect a number of their biological samples is they mail kits to these women and have them go to their doctors, and have the doctors draw the samples and then ship them to the researchers. This is a clever use of existing medical services to actually gather these data. The other thing I wanted to mention is there's a number of mobile clinic-based strategies going on: the NHANES now has huge trucks that they move around the country, that are sort of mobile labs, where they bring people into a "clinic," but the clinic itself is mobile, and you can move it around the community, and move it around the country. There's a study called HANDLS. It's actually a study being run by the National Institute on Aging's intramural branch in Baltimore at this point, and it also uses these mobiles, but that may be a technology that becomes more and more available to more of us as the technology gets to be more portable into these smaller and smaller vans.



Most recently, some of the innovations that I think are expanding our ability to get biological samples in these large, community-based studies are obviously blood spots, and I know Thom [McDade] and others will be talking about that. And now a number of assays are being developed that use saliva, the most familiar of which, to many of us, is, of course, the salivary cortisol samples, which make that something that you can track pretty easily in community-based samples.

I have actually been working with the Behavioral and Social Research Program for about the last year, and the goal of my consulting is to help them develop tools to integrate biological protocols into population-based surveys.

NIA/BSR Biological Protocols

Primary Goal: To develop a centralized resource ("reference manual") to provide information on requirements for collection of various biological specimens (e.g., blood, saliva, urine), including:

- Details of specimen collection/processing protocols
- Assay costs

Rationale: Researchers interested in incorporating assessment of biological parameters into their research could more easily evaluate the range of potential parameters along with the logistical and financial "costs" for each.

One of the things we've done is to think about developing a centralized reference manual that would provide information for people on the requirements to collect different kinds of biological specimens, including details of collecting and processing. And also, what it would cost you if you want to assay those things, which can end up being a major constraint on your ability to integrate these kinds of things. The idea here was that researchers who are interested in including these different kinds of biological measures would have a place to go to see 'What does it require?' and 'What are the logistical and feasibility issues that I need to consider?' And then, 'What am I getting myself into in terms of costs? Is this something that I can even consider in my study?' Actually, this is something I've discussed mostly with the NIA folks, and so I'd be really interested to hear back from you about whether something like this would indeed be useful. What we're trying to do is think about developing information and materials that could be available to researchers who want to integrate biological specimens in their studies.

Activities to Date

- Collected and summarized information on commonly collected biological parameters from population-based studies
  - Community-based collections
  - Clinic-based collections

What we've done so far is we've been collecting and summarizing information on commonly collected biological parameters from both community-based studies, like the ones I've been describing, and also some of the clinic-based collections, figuring that there may be people out there who have the option to get clinic-based measures, which would broaden the range of things they would be able to look at.

Chart IIA: BSR-funded Studies - Biological Protocols Used in Previous Population-Based Studies

	Mac Aging	White-hall	Taiwan	IFLS	EPESE	Hispanic EPESE	SATSA
I. HPA Axis							
1. Cortisol	X	X	X				
2. ACTH		X					
II. SNS							
1. Norepinephrine	X		X				
2. Epinephrine	X		X				
III. Cardiovascular							
1. BP	X	X	X	X	X	X	X
2. Pulse	X	X	X	X	X	X	X
3. Heart Rate Variability		X					
IV. Metabolism							
1. Total Cholesterol	X	X	X		X		X
2. HDL	X	X	X		X		X
3. Glycosylated	X	X	X				X

I have another document -- and I'm happy to provide it to anybody who's interested. This chart was meant to summarize what kind of biological information has been collected in some of the existing studies, such as the MacArthur Aging, Whitehall, and a number of other studies, most of which were funded by Richard Suzman's group. The thing to notice is some of the less technically demanding are the ones where you see x's going across. Almost all of the studies are collecting things like blood pressure and pulse, and some of the non-fasted metabolic measures.

Developing "reference manual"

- *Collected and summarized details of actual protocols*:
  - Specimen collection requirements
  - Processing methods and requirements
  - Assay Costs
- *Develop standard format for "reference manual" presentation of protocol information*

In terms of this reference manual, we've collected and summarized the details of the actual protocols, what the requirements are in terms of specimen collection, what is required in terms of processing, and what are the costs. We're developing a standard format for providing that information.

That information includes what type of sample you can get to assess for a particular biological parameter, what the collection parameters are, whether you need to have a fasting sample or not, what kind of blood tubes are needed for blood collections, whether time of day matters, how long you have until it has to get processed, whether there are any temperature constraints, what kind of processing requirements are there in terms of centrifuge, time, speed, temperature, whether there are particular shipping requirements, if you're sending these samples around, and what kind of ability to store you have, and then, as I said, the cost.

Protocol Information

- ***Type of sample*** – blood, saliva, urine
- ***Collection parameters*** - e.g., fasting, tubes, time of day, time to processing, temperature
- ***Processing*** – centrifuge, temperature, timing
- *Shipping* – requirements (dry ice, etc)?
- ***Storage*** – e.g., duration, temperature?
- ***Assay costs***

I. Type of Sample	***Glucose***
1. Serum 2. Urine 3. Sputum 4. Citrated plasma 5. EDTA plasma 6. Buccal swab 7. Other	***Serum***
II. Collection	
1. Is fasting required How long?	Fasting required ≥ 12 hrs, water only
2. Tube(s) a. amount of sample needed b. type of tube-fixative? c. # tubes if more than one	1cc red top serum tube
3. Time of day: AM PM does not matter Other	AM before 10 am
4. Is there a diurnal rhythm or other reason to impose restriction on time of day? Yes: describe:	Yes
5. Other procedures necessary after collection but before centrifuge?	Immediate inversion 8-10x room temp min 30 to max (preferred) of 60 min then 4°c min 30 to max (preferred) of 60 min before centrifugation
6. Time: Maximum allowable time between collection and processing?	

This is just an example of what the chart looks like at the moment. This is an example of serum sample for glucose, indicating that for glucose you need a fasted sample, then it goes through the various requirements for what kind of tube to use, and what the requirements are for the processing.

Next Steps #1

- Continue refinement of manual
  - Continue collection of alternative/new protocols for biological parameters already included in manual
  - Add new biological parameters and their protocols
  - Other refinements (e.g., formatting, component information)?

Some of the things we're working on currently -- and again, I'd love to get input from people in terms of whether we've covered all the things you think will be important, or are spending our time on things we shouldn't be -- we're refining this manual at this point, getting additional and new protocols to be added, and we're also working on formatting and adding other component information that people might be interested in.

Probably more interesting, I think, is we're now beginning to talk with Stacy [Lindau] and Linda [Waite] and some others about the feasibility of developing some sort of web-based distribution system for this information. Some of things we're still discussing and struggling with are, What information can be made available? We've actually gotten some lab assays from people who at this point are viewing them as more investigative and they're not particularly willing to let us put them up on a website, so we're thinking about what kind of information can be up there vs. not. How can we best provide technical support? For people who come into this and are pretty new to it, and look at the website and then have questions they want to discuss with somebody whether they could actually employ "x", "y" or "z" in their study, how might we be able to do that? One thought is to have an e-mail system that would go to some initial person who sort of triages and is able to send these messages on to experts.

Next Steps #2

- Explore feasibility of developing a web-based distribution system
  - Investigate issues relating to:
    - What information can be made available through website
    - How to best provide "technical" support
    - How to "get the word out" about website – who should be targeted in research community?

Current Biomarkers Meeting

- "Minimally invasive" biomarker technology
- Assessment of Cognitive Function
- Assessment of Sensory Function
- "Human subjects" issues relating to collection (and reporting) of biomarkers information
- "Honest Brokers"
- Summary - "What is Biomarker Collection good for?"

And then, how to get the word out that this is available, particularly to investigators who are interested in getting into the area of biomarkers, and therefore may not know about this sort of stuff.

And then I just highlighted here, for those of us who are here today, what we're hoping to cover in today's meeting, and how this will integrate into some of this work that I'm doing for Richard's [Suzman's] group. One is we're definitely interested in covering some discussion of minimally invasive biomarker technology that may be the most

promising area for those of us who want to do community-based kinds of work. Expanding our focus to more integrated kinds of assessments of functional performance, in terms of cognitive and other sensory function. A big issue is human subjects issues related to collecting and also potentially reporting back to people this kind of information which has not been part of our social surveys in the past. A discussion of honest brokers and then kind of a summary at the end of what we see as being the use of these kinds of things. So that's a little bit of the background. As Stacy [Lindau] said, my main interest here is that I am working with NIA, I am trying to develop these materials, so I'm very hopeful that the discussion today will help us move forward in developing the materials that would be of use to all of you, as well as all the others out in the community who are interested in adding biomarkers.

Lindau : Is there reaction or response to the question of whether a web-based system, a web-based users guide, one that's iterative and dynamic, would be of use to people in the room? This is something that we've been thinking about, again growing out of the NSHAP project, Jenna Mahay, Teresa Seeman, Linda Waite and I have met several times to talk about it. We want to move in that direction, but only if it can be useful. I'm seeing a lot of shaking of heads 'yes.'

Seeman: One thing we may want to try is putting some of the information we've been collecting up on the website and have people try using it to see whether you can access what you need in a way that helps you to answer your questions, because I'm guessing that our first attempts will be less than perfect.

Freidman: The other thing you might consider, associated with that, is that you can go to a listserv and you can ask questions, and you can click on "ask questions"… [audio unclear].

Seeman: That's a great idea.

Lindau: Yes, I think a listserv is a great possibility for joining us together.

Lipton: That would be a fabulous resource. All of us at NIH, the methodologic section, everybody had their own opinion on how to go assaying this or that, and how many hours fasting or before fasting, but it would be a great resource for people to simplify as many aspects of the thing. Of course, there would be a big debate, and a lot of fighting about actually who gets to write. A listserv would be a more formalized mechanism for actually making sure you had the best level of expertise or that you had x number of alternatives. Or that you weighed the measurements based on the intensity and cost and so on and so forth. That would be really fabulous.

Willis: It would be useful for studies like the Health and Retirement Study that I work for to have some citations to the literature, particularly to literature that says 'What's the relationship between this biomarker and various kinds of social, health, and academic considerations?'

Lindau: I agree, and I would add I would like to see links to literature that have validated the assays, studies that have used the assays.

Lipton: Five years ago now, a couple of physical activity epidemiologists put together an issue in the *Journal of Sports Medicine*, assessing 150 different tools for trying to measure physical activity, and the issue was obviously a hardcopy, not even a pdf file, I don't think, but it also listed all the populations in which these instruments had been validated, and the different error measurements, coefficients of variation and all that stuff. It's a huge job, what you're talking about. It seems like it would be a fabulous resource.

Hughes: This is related to Bob's [Willis'] point. One of the things also, if you really want this to be useful for people who aren't biologically trained, is to include information about what underlying process this biomarker taps. Because it's fine to look at what's already been shown, what social processes have been related to this particular biomarker, but if you're interested in developing hypotheses about what's going on in the model that Teresa [Seeman] showed, it's very helpful to know what is this measuring.

Seeman: One of the things we've been debating about is how much goes in there, and how much do you just sort of tell people, 'look you're going to have to go to another resource.' The level we can put in there won't be enough. What we've started to do is put in some brief, very non-technical definitions for some of the things, and also include definitions – what are these different kind of tubes we're talking about. I think we'll just have to have a disclaimer that says this is not meant to be comprehensive -- if you're going to do it, you have to go learn, or get a colleague to work with you. When I started getting into the biology stuff, I got a career development award, and one of my mentors was Jack Rowe. And Jack said to me, you're not going to become an endocrinologist, you just want to learn enough to talk to them. That is fundamentally what the social science types have to come to grips with. You need to also collaborate with people who know this stuff cold.

Hughes: Like having some guidance and knowing what questions to ask.

Seeman: Yes, and beginning to learn the language.

Garfield: I would ask how large a scope you want. I'm a pediatrician, so one of my first thoughts would be what's available for a kid, what can you use for a kid, and what we know about that. And while it's probably not in the purview of NIA, if you could chop off that top half, and just say, of these that we're looking at, how many have been used with kids. That would reduce the amount of work that would then be required to look at 18 and younger.

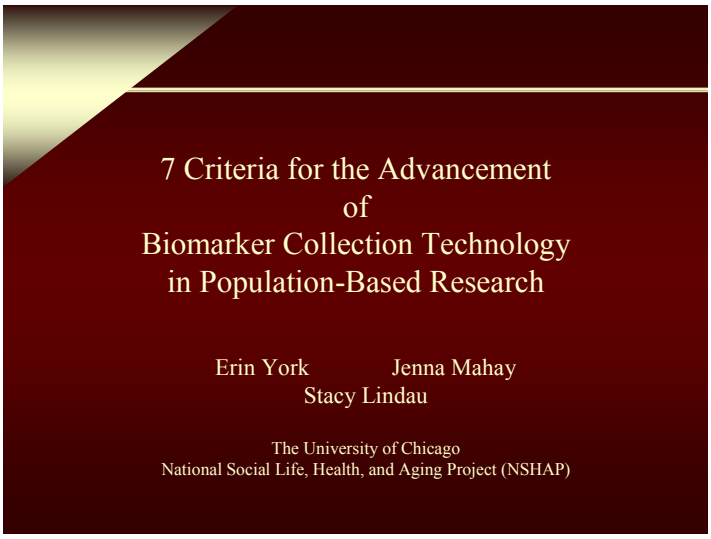
Lindau: And that relates to a point I wanted make and that we'll be talking about tomorrow. There are clear human subjects implications of a variety of assays. We're starting to think about some assays as being clinically relevant assays that need to be reported back to respondents, and other assays that are either experimental in their design, meaning they've been adapted for blood spot technology, we don't know necessarily about their relevance, or they're really truly experimental assays. We talked last year about F2 isoprostanes as a marker of dementia. Most of us agree that it's probably not responsible to report that back, and there needs to be some guidance about that. I know we're struggling, all of us, with IRBs about these issues.

McQuillan: This is really great. I've written the lab component for NHANES and we have over 400 lab tests right now. It's an enormous task, but what we might do is work together, because I can see that really beneficial, that you link to our data on the Center on Aging website where we have all our methods. And then, when people ask us -- I get a million questions from people, a lot of times they're asking about assays I know nothing about -- it would be nice to just send them to your website..

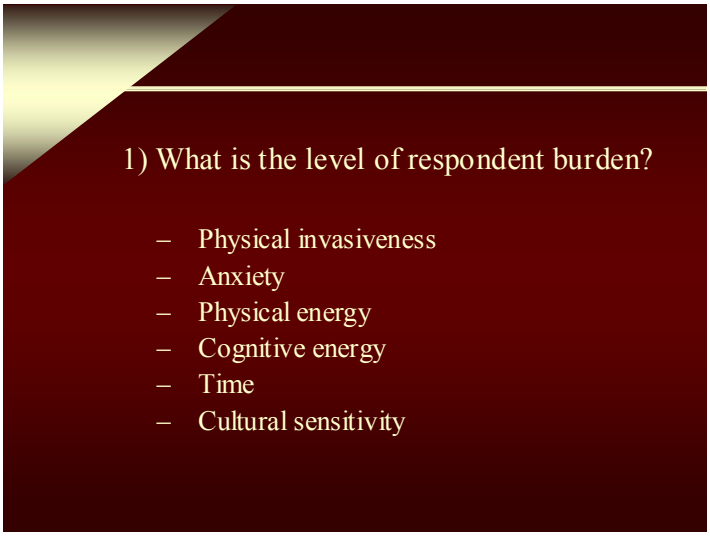
New Developments in Minimally Invasive Biomarker Collection Technology

7 Criteria for the Advancement of Biomarker Collection Technology in Population-Based Research

Speaker: Jenna Mahay

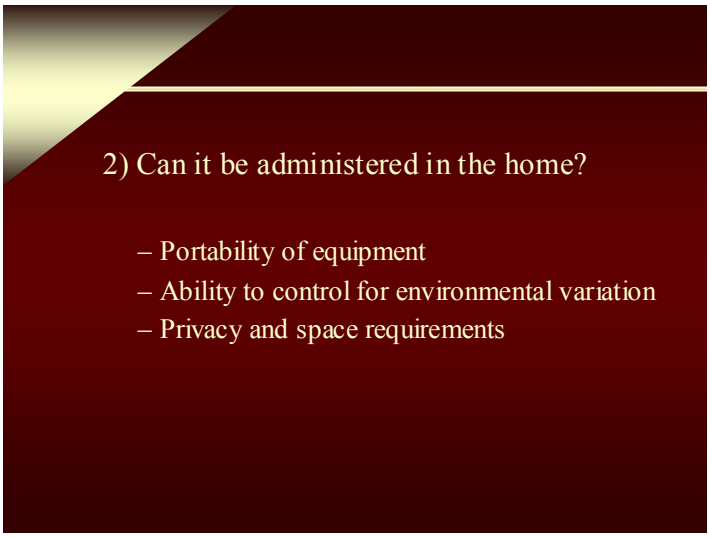


I would like to take a few minutes to talk about the seven criteria for the advancement of biomarker collection technology that we came up with in the course of developing the biomarker component of the National Social Life, Health, and Aging Project. This is a national population-based study of older adults that will be conducted in the home with non-medically trained interviewers. I'm hoping that this presentation will provide a framework for thinking about the exciting new biomarker technologies that we're going to be hearing about for the rest of the day today and tomorrow. Before I begin, I would just like to acknowledge Erin York and Stacy Lindau for their work on this as well.



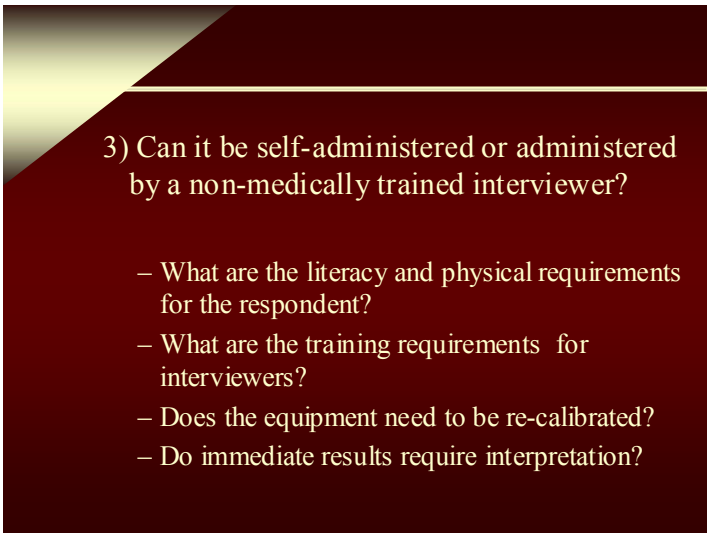
I'm going to start with the questions we had to ask when we were deciding what biomarkers to collect as part of our study. One of the most important ones was: What is the level of respondent burden? This is particularly important for population-based studies, where we need to get the highest response rate possible, and for longitudinal studies where you need to retain the largest number of respondents for follow-up waves that you can.

There are a number of components of respondent burden. First, there's the physical invasiveness. As we're going to hear from Thom McDade, the finger-stick blood spot is a much less invasive way of collecting blood than venipuncture, and will also probably get a much higher response rate because of that. Also, anxiety: How much anxiety is the procedure going to produce for the respondent?; the physical energy that it will take for the respondent to provide this biomarker; and the cognitive energy, both in terms of the comprehension of what's going to be done to them, and in terms of the mental effort needed to follow directions, if that's part of the biomarker collection. And finally, the amount of time that it's going to take. This is going to be a factor in any population-based study, where you have a limited amount of time that the interviewer has to collect the data, as well as the amount of time that the respondent is going to be able to give you.

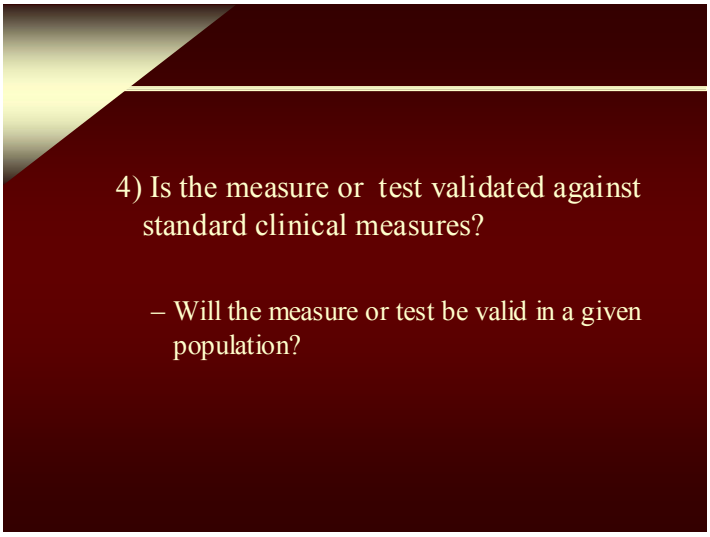


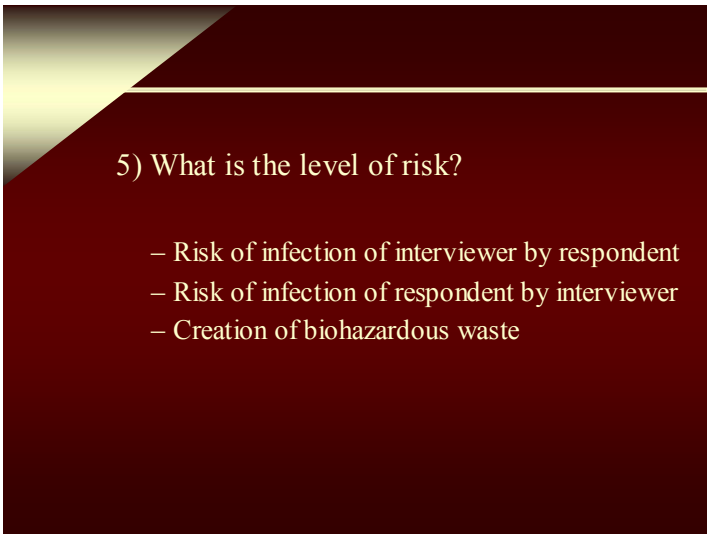
The second question was: Can it be administered in the home? This might seem fairly obvious, but there are a couple of components to this. First, the portability of the equipment, especially if the interviewer is going to be carrying a lot of equipment for different types of biomarkers with them. Also, the ability to control for environmental variation, for example, the lighting in the home if you're going to do a vision test. Is there a way to control for that? Or the room temperature, is that going to affect blood pressure? And then finally, the privacy and space requirements for the biomarker procedure.

Third, can it be self-administered, or administered by a non-medically-trained interviewer? And again, this is important in population-based surveys that are going to be done in the home. If you can have it self-administered, or administered by a non-medically-trained interviewer, that will substantially cut down on your costs. So what are the literacy and physical requirements for the respondent, if it's going to be self-administered? What are the training requirements for the interviewers, if it's going to be administered by an interviewer? Does the equipment need to be recalibrated by a professional before it is administered? And do immediate results, like blood pressure, require interpretation if the respondent asks; does the interviewer need to know critical values at which they need to seek immediate medical attention?

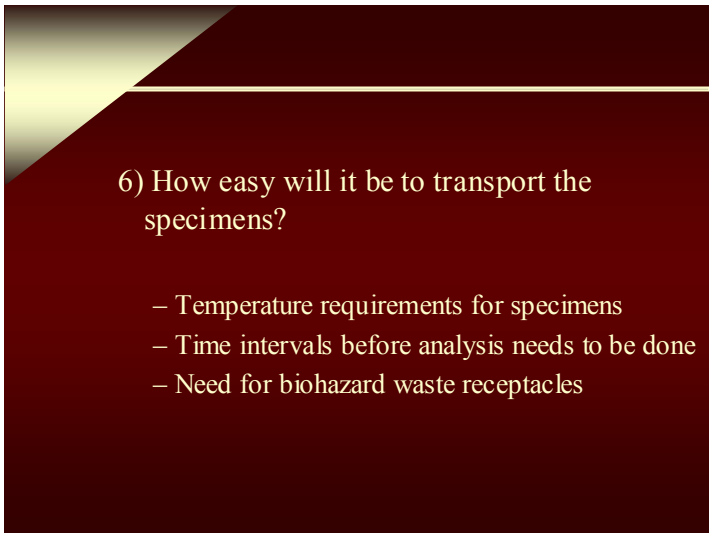


Fourth, is the test validated against standard clinical measures? For example, is salivary estrogen going to be comparable to clinically-used serum estrogen levels? And this is very important, again, in population-based research, because we can collect all the biomarkers we want, but if we don't know if they're valid against clinical measures, then what do we have? The data may not be that useful. A corollary question to this is, Will the measurement test be valid in a given population? So, for example, in an older population, such as NSHAP participants, we need to be fairly certain that the number of true positives for something like an STD test is going to be larger than the number of false positives.

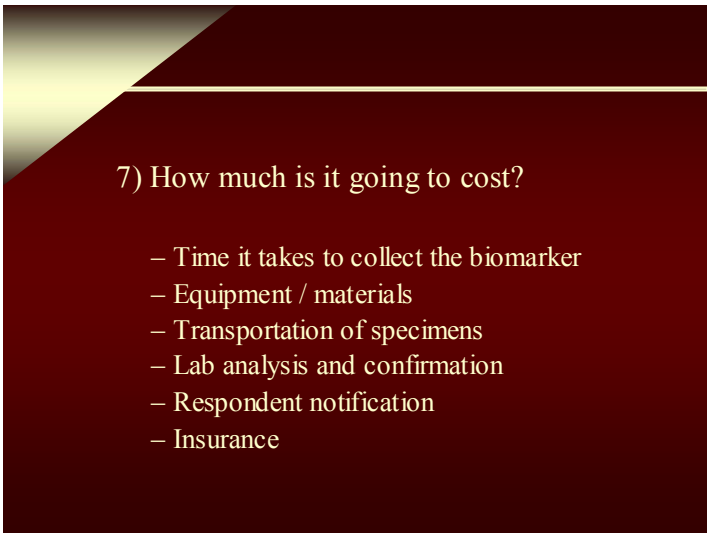




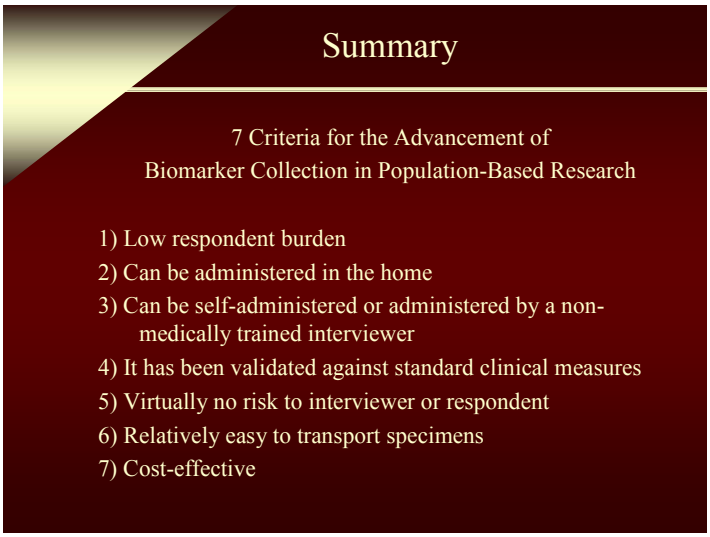
Fifth, what is the level of risk? There's the risk of infection of the interviewer by the respondent, or at least concerns about this, even if it's not an issue. There's also the risk of infection of the respondent by the interviewer, as well as the creation of potentially biohazardous waste, that people might have concerns about putting other members of the household or the community at risk.



Sixth, how easy will it be to transport the specimens? There are a couple of components of this. One is the temperature requirements for the specimen. For example, do those specimens need to be frozen right away, or refrigerated as they are transported to the lab? Other issues are: How much time do you have before the specimens need to be analyzed? And do you need biohazard waste receptacles, which can potentially be costly and bulky?



And finally, how much is it going to cost? Every project is on a budget, and there are a number of components that go into how much a biomarker can cost, which one might not think about right away. There's the time it takes to collect the biomarker within the interview. There's equipment and materials, obviously, transportation of specimens, which is going to depend on whether they need to be refrigerated, frozen, how soon you need to get the specimens to the lab, and whether they be batched and sent all together or if they need to be sent individually. There is also the costs of the lab analysis and confirmation, respondent notification, and finally, insurance.



In summary, from this analysis that we had to go through for each biomarker that we decided to collect as part of our study, we came up with seven criteria for the advancement of biomarker collection in population-based research. The first, of course, is low respondent burden. It must also be able to be administered in the home, and it must be able to be self-administered or administered by a non-medically trained interviewer, unless you have a mobile clinic or something like that. It's been validated against standard clinical measures. There's virtually no risk to the interviewer or respondent. It's relatively easy to transport specimens, and finally, it's cost effective. So I hope that you'll see this really as a working document, and I hope that we can take advantage of the collective experience in this room to add to it throughout the day, and come up with a resource that we can all use as we go forward in collecting biomarkers in our own studies.

Friedman: There's one thing that is not in there that is important to consider, which is quantity. So the blood spot, for example, may be great, but if you're going to do a longitudinal study, and you may some day decide there's some genetic component or something, you may want a lot of blood. I think that's a key question to ask yourself up front, because there's almost never enough in the end, even though a lot of it gets thrown away, when you want it, for certain things. It's a balance, obviously, burden and collection, and you have the whole storage issue as well.

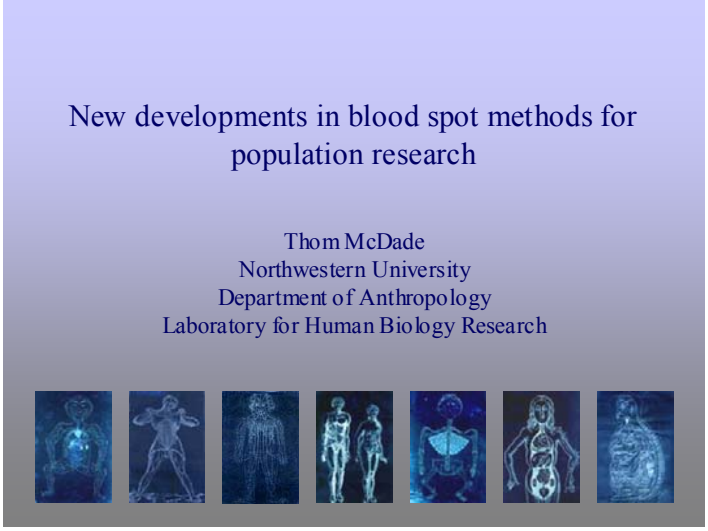
Unidentified Speaker: I just want to raise a question about the criterion about it being a validated measure, and whether or not that misses an opportunity. Many of the things, looking through last year's workshop, are things that I would consider not to be validated, but are things that there are great opportunities to be studying them.

Lipton: About the criterion that there should be virtually no risk to the interviewer or participant, I think that is going to be a real stumbling block, as the definition of risk gets much broader in every IRB and legal department that we deal with. And also, I think that the issue of the tradeoff. I mean, sometimes you do need to store a blood sample, you do need to do phlebotomy, these kinds of things, the trade-off between actually having information that you can use, is important. So I would suggest a revision: minimal risk, or risk commensurate with the value of the information.

McQuillan: That's a very good point, minimal risk. I think the NSHAP Advisory Board addressed yesterday that if they do even think about storage, we need another consent document. So that is something that has to be considered.

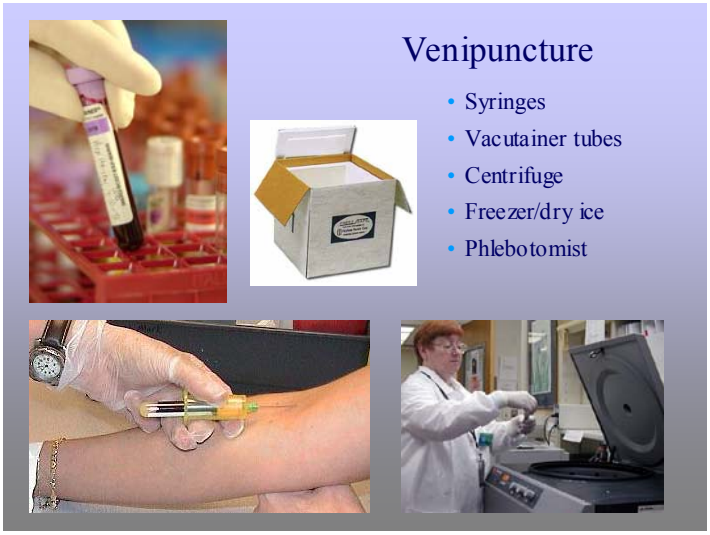
New Developments in Blood Spot Methods for Population Research

Speaker: Thomas McDade

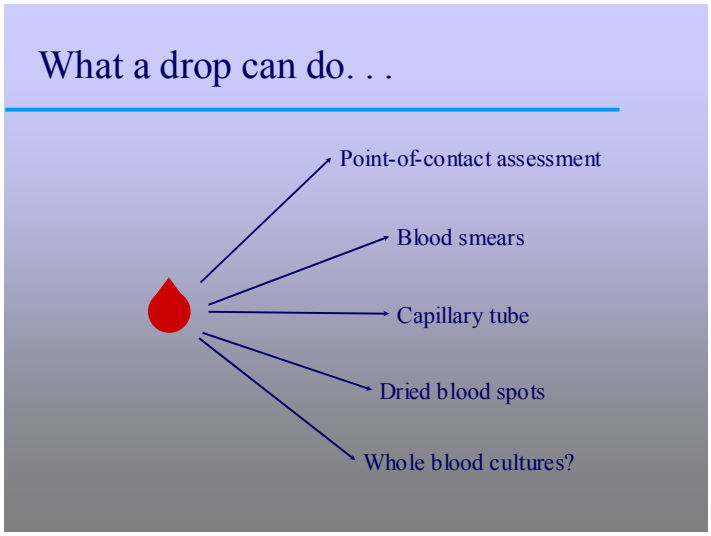


What Stacy [Lindau] and the organizers asked me to do was talk a little bit about what we can do now with blood spots, which is a method I've been using in a number of settings, and what we could think about moving towards in the future. So I'll go through what we can do now, and then end with some discussion of the future.

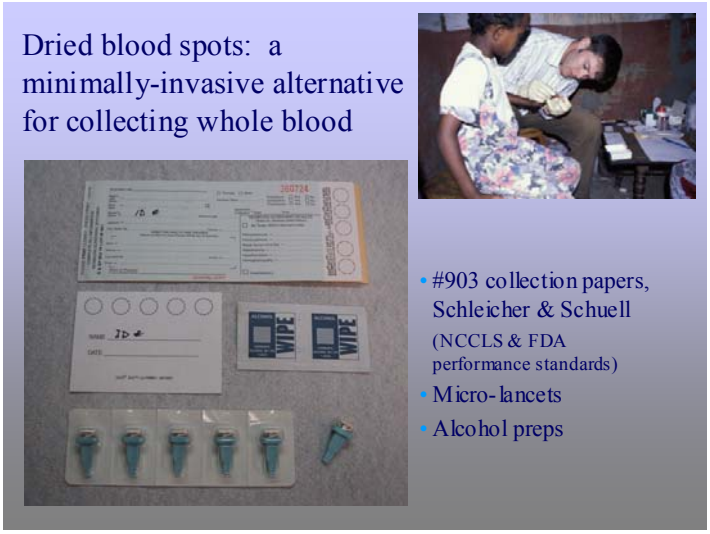
Presumably since you're here, you're interested in biomarkers, but you're probably aware that venipuncture is a major stumbling block to the community-level, population-level implementation of biomarkers, because many, if not most things of interest are in serum or plasma, and the collection of that is problematic as we all probably know from personal experience. It requires syringes, the skills of a phlebotomist, a cold chain. It is a relatively invasive, expensive, and inconvenient, to say the least, process for field-based research. So whole blood provides some alternatives, and whole blood can be collected from the prick of a finger. People who monitor their blood sugar do this all the time.



There are a number of things that can be done with just a single drop of whole blood, as new assay technologies come onboard and become increasingly sensitive. Just to put the blood spots in context, I wanted to alert you to the fact that there are a number of things you can do with a single drop of blood. There are point-of-contact assessment instruments, like a Cholestech or a Hemocue, which will measure a full lipid panel, or hemoglobin, at the point of contact with the participant. You put a drop of blood in the instrument and two minutes later you come back with the reading.

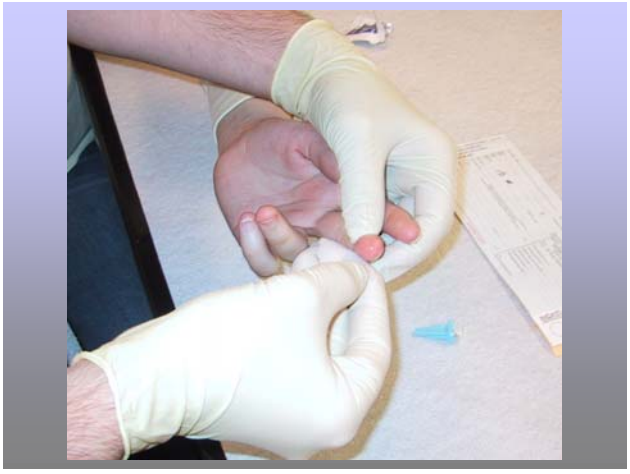


There are blood smears, thin and thick films, which have been done internationally for decades. And you can do whole blood counts, or counts of blood cells, cell blood counts and differentials. You can actually do indirect immunofluorescence, where you tag cells with antibodies and identify the ratio of CD4 to CD8 T lymphocytes, for example. You can collect blood in capillary tubes -- you can collect a couple hundred microliters in a capillary tube -- and then transport that back to the lab. This is something I actually did in Kenya and brought samples back to the U.S. for analysis. And lastly, you can put drops of blood on filter paper, and dry it, which I'll talk about in more detail. And then there are whole-blood lymphocyte cultures, which I think have a lot of potential. That's looking toward the future, and that's what I'll end with.

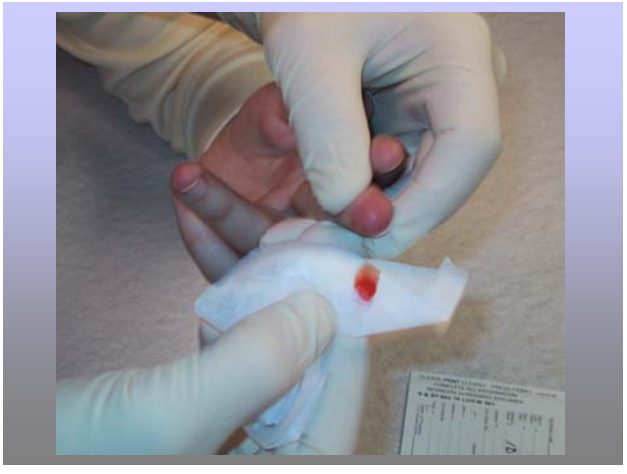


First, dried blood spots are a minimally invasive alternative for collecting and transporting blood. Everything you need is shown right here. These are the micro-lancets, and there are many different varieties that different investigators use.

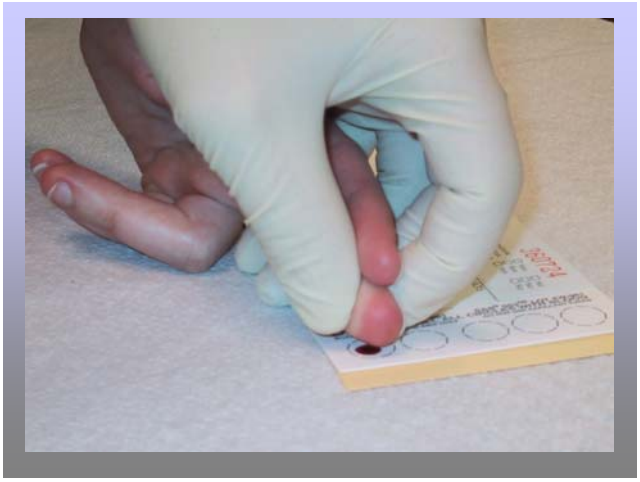
These are two different formats of the paper that are manufactured by Schleicher and Schuell. They'll customize the paper for you if you order enough of it, and what's nice about the paper matrix is that it's used in neonatal screening programs in every U.S. hospital. So it's highly regulated, highly standardized for lot-to-lot consistency and for uniformity in sample absorption. And it's regulated by the FDA and National Council on Clinical Laboratory Standards, and the CDC runs a rigorous quality-control program on each lot of papers. It's a pretty straightforward collection process, which I'll walk you through.



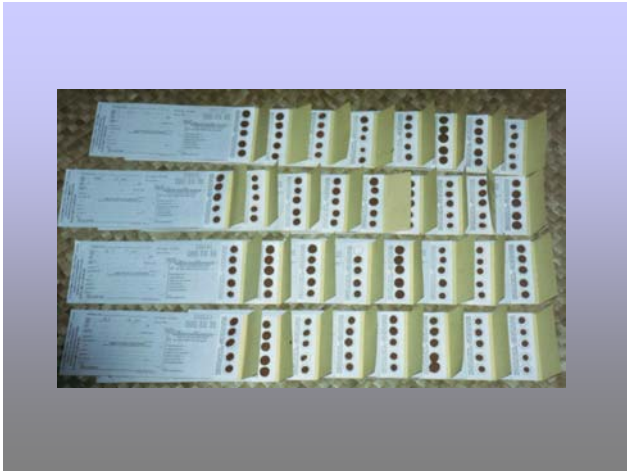
First, you wipe the participant's finger with an alcohol swab to clean it.



Prick the finger with the lancet, and wipe away the first drop of blood, which includes cellular debris and interstitial fluid.



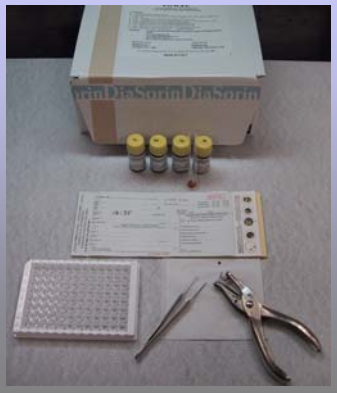
Then apply as many drops of blood to the filter paper as you can. This is an important process, and a very important potential source of error. It's not a process of blotting, it's a process of allowing a drop of blood to form on the paper, bringing the paper up to the finger, and allowing capillary action to pull the blood off. That's the only way the paper will uniformly distribute the analytes in the sample. So this is an important potential source of error. It does require some training, but it is something that non-medically trained interviewers can do in a home setting, with relatively little training compared to phlebotomy, of course.



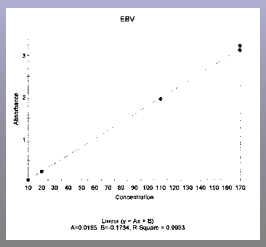
Once you collect the samples, you just lay them out to dry for at least four hours.

And then they can be stacked up, stored in a plastic bag, with some desiccant to remove any remaining moisture, and you're ready to go. Most analytes are stable in blood spots for weeks at a time at room temperature. Sample degradation happens more quickly at higher temperatures and it varies based on the analyte and the biochemistry of that analyte; this is something that needs to be evaluated before you implement this technology. But what's nice about it is that it buys you time, so these samples can be stored at room temperature in Kenya, Bolivia, Samoa, North Carolina, Chicago, wherever, and you don't have to get it to the freezer, or the centrifuge, right away.

Blood spot EBV Antibody assay



- Commercially available ELISA kit: DiaSorin, Inc.
- Assay cost: ~$5.00/sample



Once the samples are back in the lab they get stored in the freezer. When you're doing an analysis (this is for blood spot EBV antibodies which is something that I've been working with) you bring the sample out of the freezer, you take this $2.99 hole punch from Office Depot, and you punch out little discs of blood. Then you elute it in a test tube with a buffer that reconstitutes the sample. So you've basically made whole blood again. And then you run your immunoassay, and based on a standard curve, you get a quantification of the analyte in the blood spot sample.

So what can you measure in blood spots? This is from a recent paper, published by Joanne Mei at the CDC. This is a group that has been involved with the neonatal screening program, but they were invited to a session about the population-level applications of the blood-spot technology. I recommend this reference if you are interested in considering this technology. She lays out nicely some of the quality-control issues that have been implemented with the filter papers, and she has a table that lays out the things that have been measured in blood spots in humans. It's over 150 analytes, which is great, but you go through this list and you see all these things that look of interest to you, and there are no references here.

Mei et al (2001). Use of filter paper for the collection and analysis of human whole blood specimens. *Journal of Nutrition* 131:1631-6S.

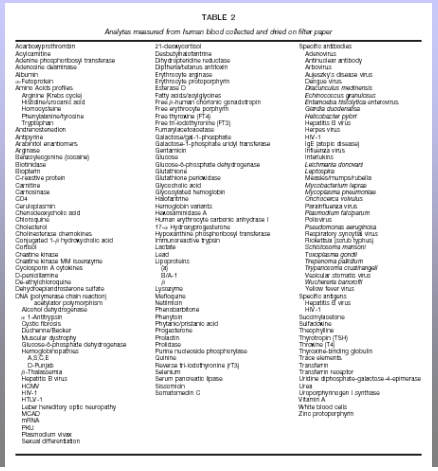


TABLE 2

Analytes measured from human blood collected and dried on filter paper

So what I've been doing in the last six months is collecting all the protocols that I can find that have been using blood spot methods. The number of protocols that have been published, and adequately validated, is probably 1/10th of this. The reason is that it requires a lot of work -- I'll talk about that in a minute.

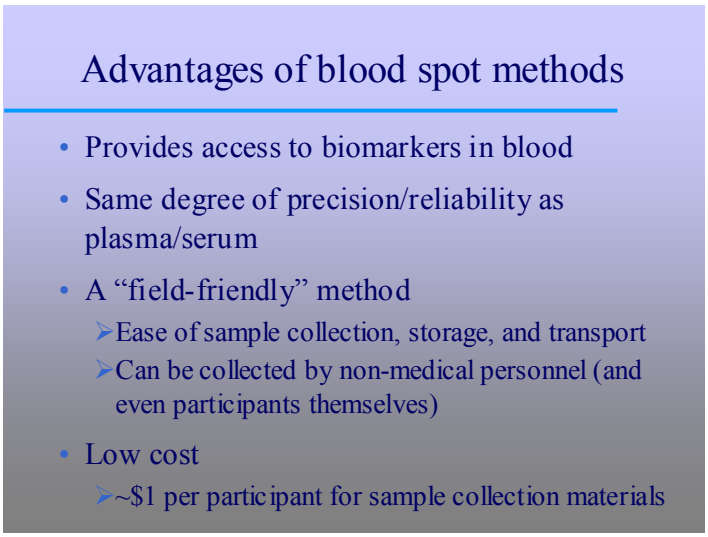
Validated blood spot assays

- EBV antibodies, CRP
- FSH, LH, estradiol, testosterone, SHBG, DHEA-S, cortisol, androstenedione
- TSH, T3, T4, leptin, HbA1c
- Hemoglobin, TfR, retinol
- Antibodies to numerous infectious agents (e.g., HIV, syphilis)
- DNA, RNA

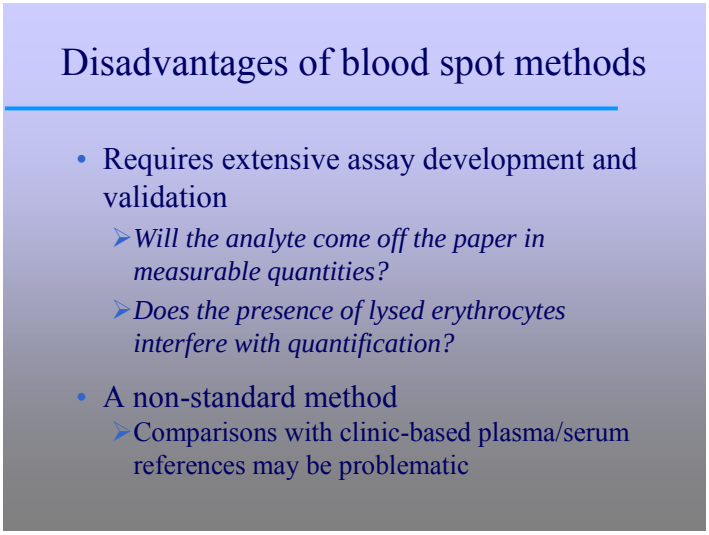
So this is a list of things that either I have measured directly, or that labs that I am familiar with have measured, or that I'm convinced the protocol has been published, the ranges are available, it's been well-validated, and I'd be comfortable saying you can measure this in dried blood-spots. There are lots of things that are not on this list that probably can be measured, but to make me comfortable, they'd require some additional development and validation. There are measures of immune function and inflammation, these are two methods that I've developed; measures of reproductive function and stress hormone activity; measures of metabolic function,

measures of nutritional status, iron status, retinol, and antibodies against numerous infectious agents. Antibodies are very robust in bloodspots and easy to bring out of the paper. And then there are lots of applications for DNA and RNA, which is not something I've done personally, but a lot of people have used this method for that.

So what are some of the advantages here? One of the obvious ones is that this method provides access to biomarkers in blood that would not otherwise be available to you in a population-level setting. And, if done correctly, in terms of the collection and transport of the sample, and the validation of the assay, you can get the same degree of precision and reliability in a bloodspot assay that you can get with a plasma or serum assay. And it's a field-friendly method in that samples are relatively easy to collect, easy to store, easy to transport, and non-medically-trained personnel can collect the samples. Even participants themselves can collect the samples. There have been a number of studies, mostly in terms of validation of home-based HIV testing kits, that have shown that participants are very capable of collecting samples themselves. And the cost is pretty low. It's about a dollar a sample for the materials, the paper, the lancet, the alcohol prep, and a pair of gloves.



It's also very important to be clear about the limitations here, and the major disadvantage, as I see it, is that it requires extensive assay development and validation, and if you want to do what Jenna [Mahay] is saying, collect biomarkers that have external validity and comparability to what we know about biology in clinical settings, we need to do our homework ahead of time before we can apply these methods.



There are two major issues here that are non-starters if you can't get past them. The first is, will the analyte come off the paper in measurable quantities? For some things, they're just not present in enough quantity to come out of the paper, or for some analytes the paper doesn't release the analyte, or the process of drying and being exposed to air on the paper oxidizes the analyte. Fortunately, that does not happen to many analytes, but this is an issue that needs to be evaluated.

The second is, does the presence of lysed erythrocytes, which happens when the blood sample dries, interfere with quantification of the analyte you're interested in? Again, this is not an issue for most things, but I was interested in developing a ferritin assay a few years ago as a measure of iron status. Red blood cells contain ferritin, which is released upon lysis, making quantification virtually impossible. So those are two major issues. Once you get past those potential hurdles, then you start looking at,

what's the sensitivity here? What's the precision, the reliability, the accuracy? You do the things you'd do with any other serum or saliva-based protocol to validate your assay.

The second major disadvantage -- which some people see as a major disadvantage, I don't -- is that it's a non-standard method. Whole blood, particularly in dried blood spots, is a non-standard diagnostic fluid where the clinical standard is obviously serum or plasma. And if you want to make explicit comparisons based on clinical references that are based in plasma or serum, you can have some difficulty here. There are some ways to get around that, in that you can do matched blood-spot/plasma samples and do analyses of both, and get a regression, and basically apply a correction factor and get what would be a plasma equivalent. That's been done in previous research and the correlation between bloodspot and plasma methods tends to be on the order of R~0.95; the correlation tends to be very high. So you can do this reasonably, but this is something you need to be aware of if you're talking to a clinical crowd that's expecting to see plasma and serum, because they're not going to like this. So that's where we're at right now with blood spots.

Future directions

- Develop and validate additional blood spot assays
- Investigate the potential of new multiplex immunoassay technology
- Explore the feasibility of field-based "challenge" protocols

Thinking about where we can go in the future, the first is just basically continuing to do what some of us are doing, which is to develop and validate new blood spot assays using existing assay platforms, ELISA, RIA, FIA, and this will be guided by science. What are new biomarkers of interest, of relevance to social, cultural, psychological processes that we're interested in? Ask what biomarkers we want to measure in community-based settings, and develop an assay for those markers.

The next is to investigate the potential of some new assay platforms. Multiplex immunoassay technology allows us to simultaneously measure multiple analytes in the same sample. This overcomes one of the limitations of bloodspots, in that you're collecting a very small quantity of sample. If you can measure more things in the same quantity of sample, then that overcomes one of the limitations here.

Multiplex immunoassay

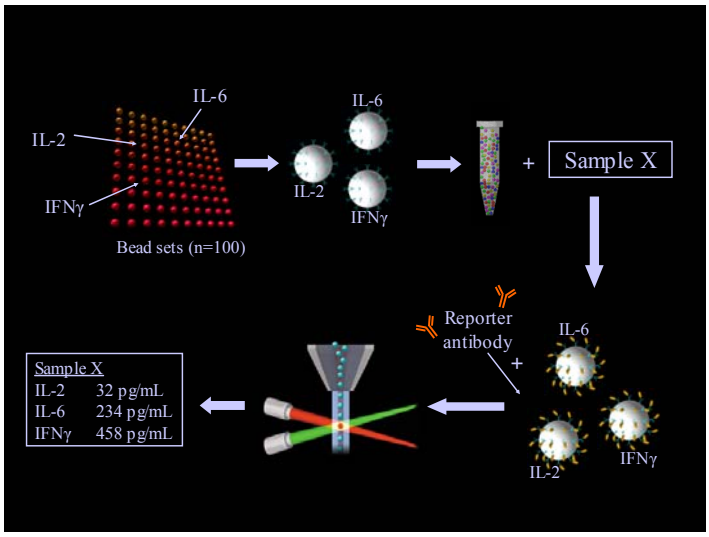
- Particle-based flow cytometry
- Simultaneous assessment of up to 100 analytes in a single sample
- Saves time, money, and *sample*

Luminex Multi-analyte flow analyzer

And the third is, can we explore the feasibility of some field-based "challenge" protocols? A challenge protocol is where we activate the biological system. They give us a lot of very interesting, important information, but they require a certain degree of control over the sample that is usually not amenable to field-based conditions. But with some of these new methods, I think we can start to question the assumption that we can't do challenge protocols in population-level field research. So I'm going to just suggest a possibility for that.

So what is this new technology? This is the instrument, a Luminex Multi-analyte flow analyzer. It's been around for five or six years, and is revolutionizing immunoassay research. How it works is that instead of using a microplate, which any of you lab geeks would know is what you bind antibodies to and you add your sample, and you do all this stuff to quantify it. The Luminex replaces the microplate with polystyrene microspheres, and there are a hundred different sets of microspheres, each of which has a unique fluorescent signature.

What you do is you bind your interferon gamma antibody to bead set 57, say, and your IL-6 antibody to bead set 25, and then you incubate it. You mix all your beads together, you incubate it with your sample, so your IL-6 binds to one bead set, the interferon gamma binds to another, the IL-2 binds to another, and then you incubate it with a Reporter antibody. That then makes a little sandwich with the IL-6 and interferon gamma 2 in the middle, and then you send it through the instrument, which, through a convergence of lasers and sensors identifies what bead this is, says, 'okay, this is the IL-6 bead.' And then another laser lights up the reporter antibodies, and says 'okay this is IL-6, how much IL-6 is there?' And so it gives you a quantification for each of these analytes. That's the hocus pocus part.



Multiplex immunoassay

- Particle-based flow cytometry
- Simultaneous assessment of up to 100 analytes in a single sample
- Saves time, money, and *sample*

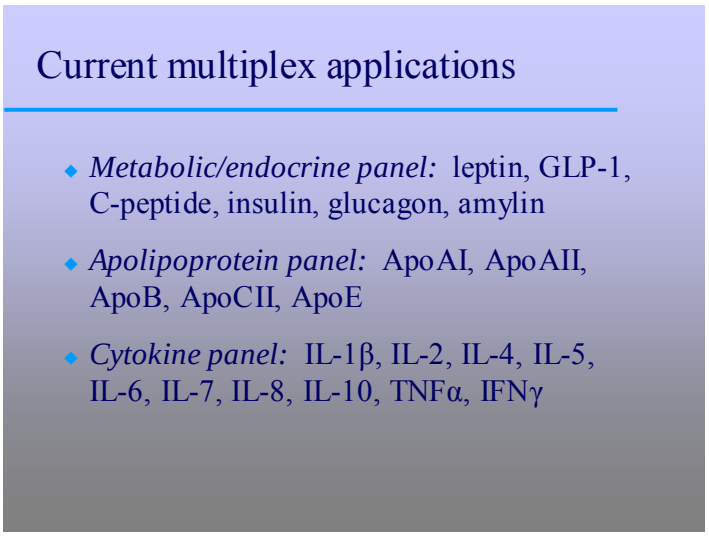
Luminex Multi-analyte flow analyzer

Analysis of 10 cytokines requires:

Method	*Time*	*Cost*	*Volume of Sample*
ELISA	40 hours	$115/sample	2,800uL
Multiplex	4 hours	$17.50/sample	25uL

As a demonstration of what this buys you, let's say you want to analyze ten cytokines -- this technology has really been developed and applied in analysis of cytokines, because there's been this realization that a single cytokine is not that meaningful. It is panels of cytokine expression that are most meaningful, so measuring multiple cytokines simultaneously is a good thing to do.

In the old way, the ELISA way, if you wanted to analyze ten cytokines, that would require 10 assays, each of which takes about 4 hours, so that's about 40 hours of technician time, and a total cost of about $115 a sample. And in terms of quantity of sample, you're talking about almost 3 milliliters of sample. With the multiplex, you can do all that in one assay, in four hours, for less than $20 a sample. And it will only cost you 25 microliters of sample. So you begin to see why I was very excited about this technology when I saw it come on the market, particularly given its potential applications for blood spot assays.



So in terms of current applications, Linco Research makes a panel, what they're calling a metabolic endocrine panel, where you can measure leptin, GLP-1, C-peptide, etc., all these things simultaneously. In 25 microliters of plasma, they also make an apolipoprotein panel where you can measure all these lipoproteins simultaneously. And there are many manufacturers of cytokine panels. I just picked the major cytokines of interest to me and many investigators, but there are many others that can be measured as well. And, the nice thing about this platform is that it's an open platform, so you can buy the polystyrene microspheres from the manufacturer, and coat them with your own antibodies, so you could come up with, for example, an allostatic load panel.

[Question from audience – audio unclear]

McDade: One drop of blood is the equivalent of about 50 microliters of whole blood. And then in terms of quantity of sample and of storage, a filter paper has five circles on it. If you do it right and you fill each of those circles with about fifty microliters, you have 250 microliters of sample. The EBV and the CRP assays that I've developed each require one punch, one 1/8th disc of blood. You can get seven of those discs out of one drop. So seven times five is 35 punches out of a single card. So EBV and CRP would only require two of those. What I'm talking about here might require five or six or seven punches. It's going to require more sample, but the payoff there is potentially greater.

So can we think about applying challenge protocols to population-based research? Here we have our very robust participant who's not going to be scared by a little finger prick, and particularly for my interests in social, cultural processes and stress, and how they affect immune function, I'm interested in patterns of cytokine production. And one of the things with cytokines is they're much less interesting if you just measure them, much more interesting if you can stimulate the system and look at patterns of expression. And I'd all but written that off for population-based research, but with some of this Luminex technology coming online, I started thinking, well, maybe there's a way to implement some of this stuff. So we can think of a couple different types of challenge paradigms that might be amenable to the kind of research that we're interested in. One is a simple vaccine challenge, where we give our participant a vaccine. This is something we've done in the Philippines, which was really seen by participants as a benefit for participating in the research, because we gave them something that was important locally, a typhoid vaccine. And then we measured two weeks later the antibody response, the vaccine titers before and after the vaccine. This is a nice, functional measure of immunocompetence. But we can also think about adding a cytokine profile to that.

The other potential mechanism, which is something that's been used by psychoneuroimmunologists and lab-based researchers for quite some time, is to take a sample of blood and pulse it with a mitogen like PHA, to stimulate the lymphocytes, get them active, and see how fast they replicate, or see what type of replication, or what patterns of cytokines they produce. So potentially you could draw a finger prick of blood, put a drop of it in a lymphocyte culture with the right medium, and then 48 hours later, analyze the pattern of cytokine expression. It's probably not going to happen in samples of five or ten thousand people, but it might happen in the future in some samples we're interested in, and it's going to give us some very interesting insights into physiology that we haven't really thought about before. So this is a few years down the road. I think we're getting some money from NIA to play around with this stuff next year, but I just raise this to get us to think beyond just getting a simple sample and measuring what's clinically relevant right now, but thinking about actually challenging the physiology and pushing things a little bit forward.

Wolfson: You're saying that these things have to air dry for four hours. So if you have an interviewer going into the home and doing this, does that have to sit out for four hours before they go off to the next person?

McDade: No, it doesn't. When the sample is liquid, it's vulnerable to degradation. So the faster it dries, the better. What I've done in the past is collect the sample and then leave it out to dry for as long as I could. So if I was finishing an interview or doing whatever, leave it out to dry for as long as possible. The papers come with a flap, and then you close the flap and put it in your plastic bag, and then go wherever you need to go, and then as soon as you get there, open it up again and allow it to finish drying.

Garfield: Two questions. One is room temperature, is that going to decrease peripheral blood flow to a point where you're actually going to decrease the amount of blood that you can get? And the other is, I know in the NICU the fingers are too small, so they use heel sticks. Have other parts of the body been used that may produce more blood, like a heel?

McDade: Well, certainly with infants the heel has been the way to go, and I would say that for infants that there would be no problem in doing it. Some researchers assessing hemoglobin have actually used the earlobe as a place to collect the blood. I've never done that.

Room temperature seems to be fine. When it gets below room temperature, then you can get some problems with peripheral blood flow. So there are a couple ways around that. One is to bring heating pads and have the participant hold those for a few minutes, or have them put their arm in warm water, or you can just get them to do this kind of thing, to throw some blood in there. And it helps if you have their arm below their knees, and that gets blood flow down there, too. These are the kind of tricks you teach in training to get the blood flow going, but that's an issue.

Mullan Harris: For the use of blood spots for DNA, do you know what the quantities would be for a genome scan?

McQuillan: We just did that in a pilot. It's very low. We actually had to drop it. It's so low that if you want to do more than a couple of genes and snips, it basically was – I can give the summary of this, but it's something that I thought was going to be great and it didn't work out. You just get too little blood to do this.

I just have one quick question. If you have to do a regression in order to correlate this with plasma or serum, how do you get reported findings to individuals on this? Is this something that the University of Chicago's [NSHAP study] is going to have to do and have programmed?

McDade: The regression would give you a plasma equivalent. So if I were to report, which I would try to avoid doing, then I would give this to the person in the plasma equivalent. Because that's what they're going to see if they talk to their physician or if they read in Newsweek what the guideline is for CRP.

McQuillan: So you're going to have to actually incorporate that in your study.

Lindau: You can either do that, or we say that these are experimental assays and not report them. This is part of what the discussion needs to be tomorrow.

Fendrich: If these have been done in population-based studies, what is the data on response rate to this kind of procedure?

McDade: It's quite good. Internationally, where I've done this with three populations now, I've had two or three people refuse. That's not ever been a problem. In the United States, to date this has been applied in a number of settings. In a way most similar to what most people are interested in here, in rural North Carolina with 9, 11, and 13 year olds, and the acceptance rate there was over 90%.

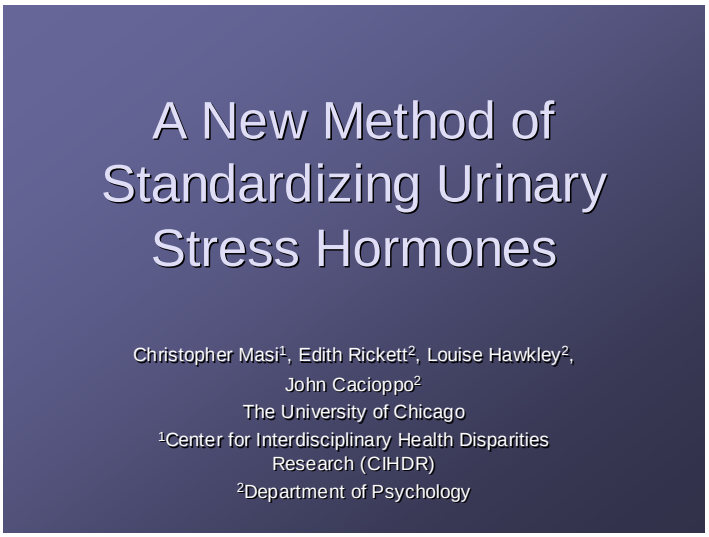
Friedman: Two quick comments. One is I am on the Hopkins IRB, and I have been for several years. I'm not going to be in tomorrow, so I just want to say that my impression, and I think you need to argue with your IRBs if they feel otherwise, is that if this is an experimental assay and not a clinical assay, you have absolutely no requirement to report them to the patients. I know that's a concern to people, but I think it's just misinterpreted by many IRBs.

The second thing is the buccal swabs for DNA. Because you do need fairly high quantities (we're using them in a study I'm doing and I work with a guy who works very closely with them, has used them for years), they're coming up with a newer amplification technology. In fact, he's recommended to me that you leave it unprocessed on the nylon swab. In the past they've been saying that you should really elute it out, and then have the DNA separate, but with the newer technology that he thinks is really close, and will give you a lot of DNA, you don't want to have done that. And in that case, it can just

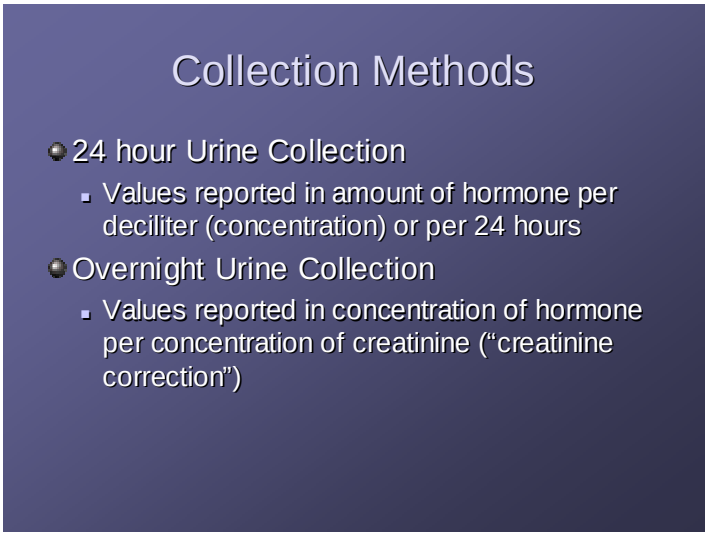
sit around for a really long time. You can leave it in the office for a couple weeks, and then bring it over to the freezer. A lot of people seem to know that, but I just wanted to point that out because I don't think this [blood spots] is going to be a way for genetic testing; you need a fair amount of DNA to do this.

A New Method of Standardizing Urinary Stress Hormones

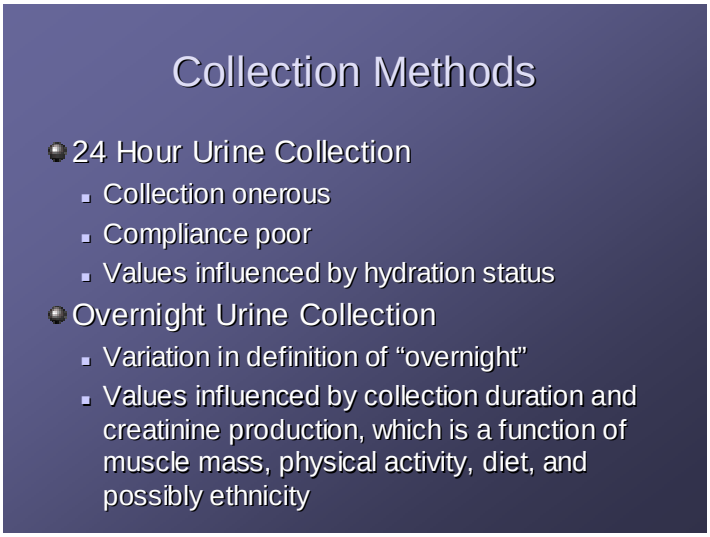
Speaker: Christopher Masi



I will discuss a new method for standardizing urinary stress hormones. I'm not going to give a comprehensive overview of stress hormones, but I will discuss a technique that was developed in collaboration with John Cacioppo at the University of Chicago within his program pilot grant looking at health, aging, and social relations among older individuals.



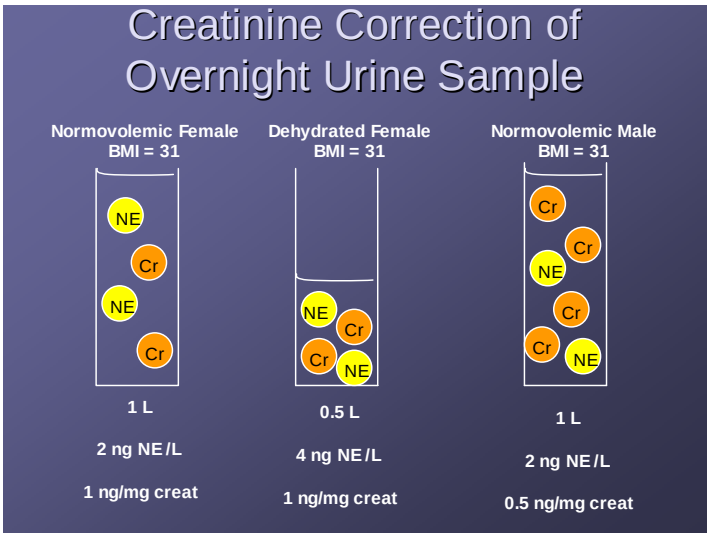
Currently, there are two primary methods of obtaining urine for hormone assays. The first is the 24-hour urine collection, where values are reported in amount of hormone per volume of urine. This is a simple concentration. The other method of obtaining urine is the overnight collection method, where values are reported in amount of hormone per volume of urine divided by amount of creatinine per volume of urine. This is a urinary hormone ratio.



Unfortunately, both of these methods have major drawbacks. For example, the 24-hour urine collection is onerous. A lot of people don't like to carry the containers around for 24 hours and compliance is therefore poor. In addition, hormone concentrations obtained through this method can be influenced by hydration status. Hormone concentration will be low if a lot of urine was produced over 24 hours while concentration will be high if the study subject is dehydrated and produces little urine over 24 hours.

With the overnight urinary collection, there is variation in the definition of 'overnight.' Study subjects are typically instructed to collect all urine after they go to bed for the evening. This includes any voiding during the night as well as the first morning void. But people go to bed at different times, and across studies there are different criteria for when urine is first collected.

But I think a bigger issue, and we're seeing this acknowledged in the literature, is that urinary hormone ratios are influenced by creatinine production. Because hormone ratios are reported as concentration of urinary hormone over concentration of urinary creatinine, higher urinary creatinine concentration in the denominator leads to lower hormone ratios.



This slide illustrates this phenomenon. Here are three urine samples: one from a normovolemic female, one from a dehydrated female, and one from a normovolemic male. If one liter of urine is produced by the normovolemic female and if we report urinary norepinephrine as a simple concentration, then we get a value of two nanograms of norepinephrine per liter. In the dehydrated individual, with a smaller overnight urine volume, .5 liters, the simple concentration is 4 nanograms per liter. Clearly, concentration is not the way to report hormone values in overnight urine samples. Traditionally, this phenomenon has been adjusted for by assuming that all individuals excrete creatinine at approximately the same rate: 1 gram per 24 hours. If we divide the urinary hormone concentration by the urinary creatinine concentration for both of the samples from women, then we find that both samples have equal urinary hormone ratios despite the differences in volume. The hormone ratio for both is 1 ng of norepinephrine per mg of creatinine. One problem with this approach is that males excrete more creatinine per day than females, and we're finding that some racial/ethnic groups excrete more creatinine than other racial/ethnic groups. And so, if you're looking at a normovolemic male, who excretes a liter of urine, two nanograms of norepinephrine per liter here, the hormone ratio goes down to .5 nanograms of norepinephrine per milligram of creatinine because there is more creatinine in the urine sample. Does this mean this male produced half the norepinephrine overnight compared to the females? No, the low hormone ratio is an

artifact of the greater amount of creatinine excreted. So this issue of differential creatinine excretion is something we and others are struggling with.

Effect of Creatinine Correction

	Gender	Mean
Nanomoles/24 h		
Norepinephrine	M	262*
	F	220
Epinephrine	M	41*
	F	27
Nanomoles/g creat.		
Norepinephrine	M	200*
	F	230
Epinephrine	M	31
	F	28

Gerlo et al. Clin Chem 1991;37(6):875-878 *p < 0.001

Here is a study which demonstrates this phenomenon. This study collected 24-hour urine samples and reported hormone values both as hormone per 24 hours and hormone per gram of creatinine. You can see that when the values are reported as hormone per 24 hours, males are found to excrete more norepinephrine and more epinephrine compared to females. However, when creatinine is placed in the denominator, it appears females excrete more norepinephrine per 24 hours.

Is there a better way to account for muscle mass and urine concentration in overnight urine samples?

So the question is, "Is there a better way to account for muscle mass and urine concentration in overnight urine samples?"

Led by Dr. John Cacioppo, CHASRS is a population-based study of 81 blacks, 66 Hispanics, and 83 whites, all from Cook County and all aged 50 to 67.

Methods

- Chicago Health, Aging, and Social Relations Study (CHASRS)
- Study Population (n = 229)
  - 81 Black (44 female, 37 male)
  - 66 Hispanic (33 female, 33 male)
  - 83 White (43 female, 40 male)
  - Cook County
  - Aged 50-67

Methods

- Overnight Urine Sample
- HPLC assay for epinephrine, norepinephrine, and cortisol
- I. Volume Correction
  - ng hormone/dL volume
- II. Creatinine Correction
  - ng hormone/mg creatinine
- III. Residualized Creatinine Correction

We instructed study participants to void before going to bed. Once they had retired for the evening, participants were instructed to collect all urine voided through the night as well as the first morning void. We then performed assays for epinephrine, norepinephrine, and cortisol and we analyzed results using volume correction, standard creatinine correction, and a new method which we call residualized creatinine correction.

To perform residualized creatinine correction, we first needed to measure fat free mass. Because fat free mass is a reflection of muscle mass, obtaining values for each study participant allowed us to account for differences in creatinine excretion which were due to differences in muscle mass.

Residualized Creatinine Correction

- Fat Free Mass
  - Bioelectrical impedance analysis
- Partial Regression Analysis
  - First calculate residualized scores for creatinine levels to account for differences in muscle mass
  - Then calculate residualized scores for urinary hormone to account for urine concentration

Results

	White	Black	Hispanic	ANOVA
Age	58.2	58.2	55.6	p<.001
Education				
< H.S.	6	12	13	**Overall**
H.S. Grad	18	28	26	X^2
> H.S.	58	40	26	p<.05
Income				
< $20,000	6	19	9	**Overall**
20k – 75k	37	40	38	X^2
> $75,000	37	16	14	p = .121

We did this by performing a partial regression analysis, where we calculated residualized scores for creatinine. This allows us to have a value of urinary creatinine for each individual which is corrected for the individual's fat free mass. This essentially removes the male/female (and possibly the ethnic) differences in creatinine production. I will explain this in greater detail in a minute.

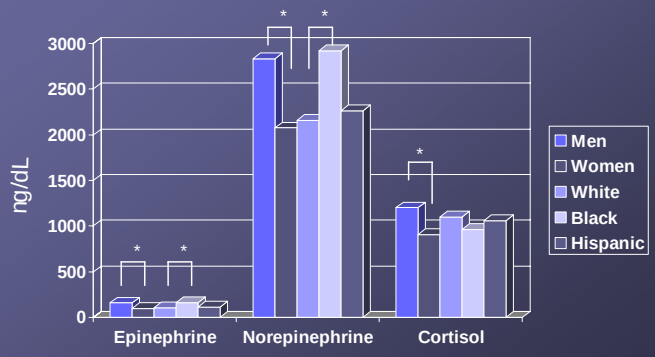
Body Size and Urinary Creatinine Concentration

	Men	Women	White	Black	Hispanic
Weight (kg)	92.5	81.3*	84.6	92*	82.6
BMI (m^2/kg)	31.2	31.8	29.9	33.2*	31.4
Creat (mg/dL)	114.2	67.23*	82.24	103.2*	84.05
FFM	69.8	44.0*	56.7	57.2	54.2

* p <.05

In our sample, Hispanics were slightly younger than whites. Education also differed to some extent, with whites having more advanced education. Income, overall, did not differ in study group.

I. Volume Correction

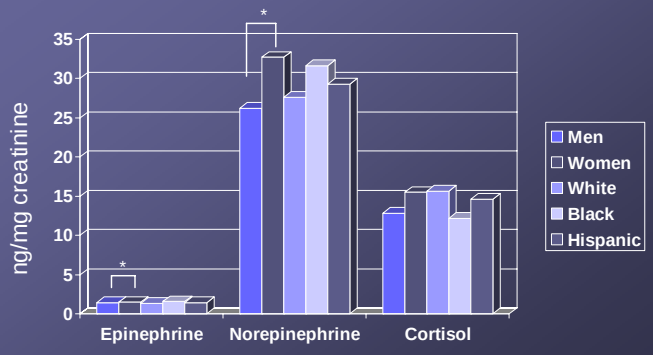


* P < 0.05

Here we show some of the gender and ethnic differences that have also been reported in other studies. We found a difference in weight by gender, and also by ethnicity. Body mass index was similar by gender but higher among blacks compared to whites. As you can see, there's a much higher urinary creatinine concentration among males compared to females. We also found what the literature has shown regarding higher urinary creatinine among blacks compared to whites and Hispanics. Here is our fat free mass measurement, indicating males have higher fat free mass. We didn't see significant differences by ethnicity in fat free mass.

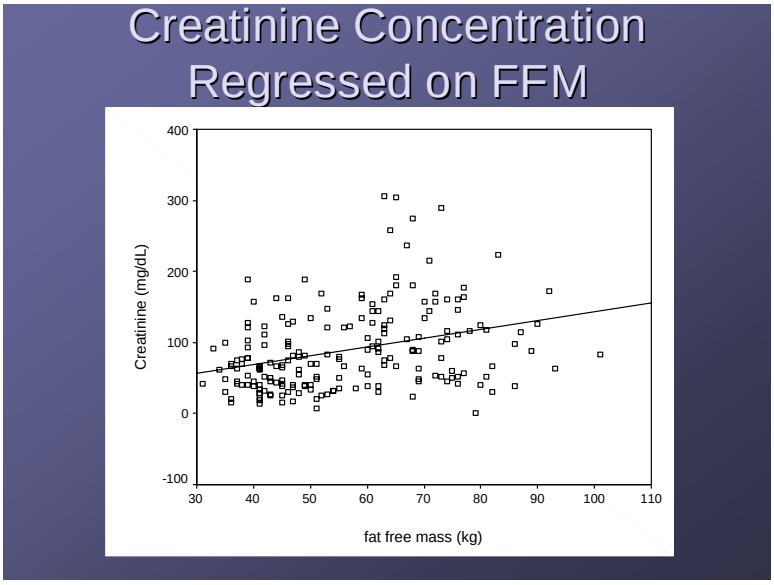
Here are the results for our volume correction analysis. Because there is so much variability in urine production in overnight samples, this type of correction is usually only conducted when there are 24-hour urine samples. But we conducted this analysis as a comparison and we found that males excrete more epinephrine than females. We also found blacks excrete slightly higher epinephrine compared to whites. Similar patterns were found with norepinephrine, with males and blacks excreting more compared to females and whites. For cortisol, we found males excrete a little bit more than females in this volume correction analysis.

II. Creatinine Correction

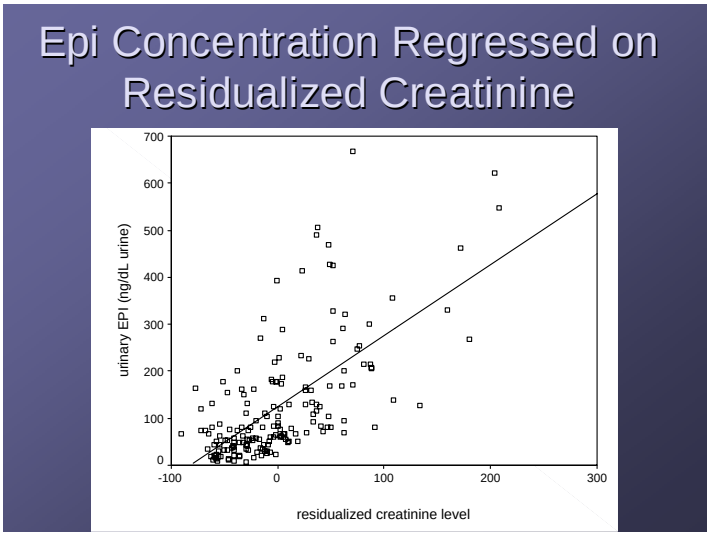


* P < 0.05

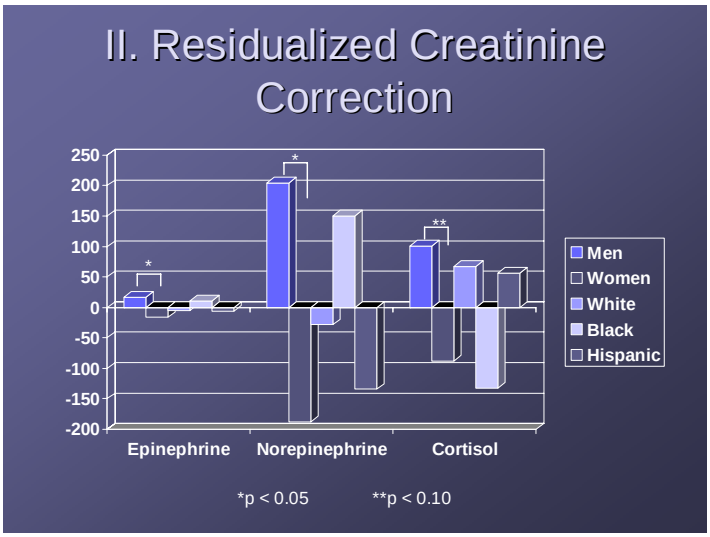
In the standard correction used with overnight urine samples, i.e. creatinine correction, we found a reverse profile of hormone excretion. That is, here we found that females excreted more epinephrine and norepinephrine compared to males. We did not find ethnic differences in hormone excretion using this standard creatinine correction.



In the two-step approach to this residualized hormone analysis, the first step is to regress creatinine on fat free mass and develop a prediction line. What can then be calculated is a residual creatinine score for each study participant. For each one of these boxes, that score is the distance from the box to the regression line, and this distance represents the amount of creatinine in the urine not due to muscle mass. So what exactly does this residual creatinine score represent? We believe this value reflects the true urinary concentration. A very dehydrated individual would tend to have a higher residual urinary creatinine value and a more concentrated urine while a lower residual urinary creatinine value would be from someone who's very hydrated and has more dilute urine. The residual creatinine score can therefore be used as a way to account for urinary concentration without being confounded by muscle mass-related differences in creatinine production.

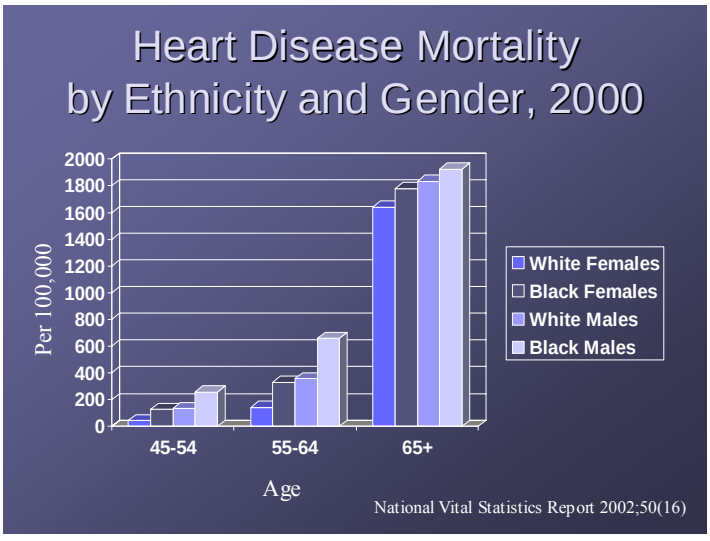


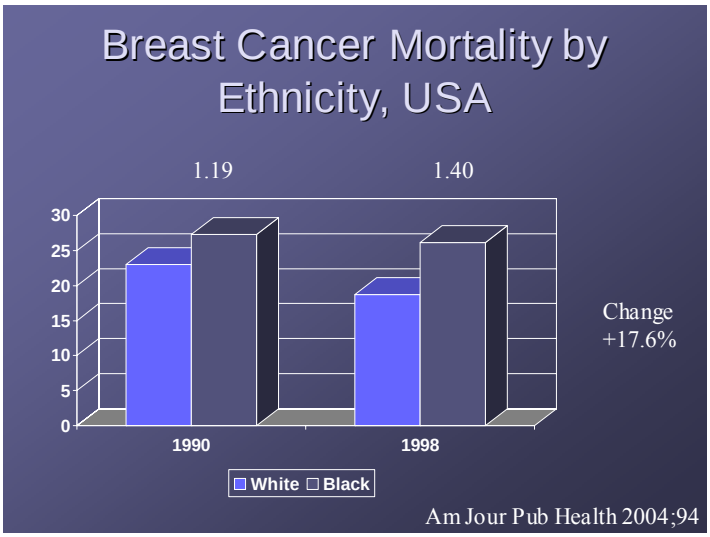
The second step is to regress the urinary hormone on this residualized creatinine values. This yields a residualized hormone level, and this is the amount of hormone which is independent of the residualized creatinine level. We feel that this gives a better reflection of the amount of hormone produced overnight. Since we're regressing the hormone concentration on the residualized creatinine value, we are accounting for the concentration of the urine.



With this approach, we find a pattern that is similar to the volume correction approach that is typically used for 24-hour urine samples. For example, we show that men excrete more overnight epinephrine than women, and we also found men excrete more overnight norepinephrine than women. The gender difference in overnight cortisol excretion did not quite reach statistical significance. We did not find statistically significant differences by race/ethnicity in epinephrine, norepinephrine, or cortisol. But we found some intriguing results which suggest, and maybe with a larger sample size we could sort this out, that perhaps blacks excrete more epinephrine compared to whites and Hispanics, and perhaps lower levels of cortisol. But we cannot make this conclusion based upon these data. I want to spend a couple of slides discussing the implications of this study.

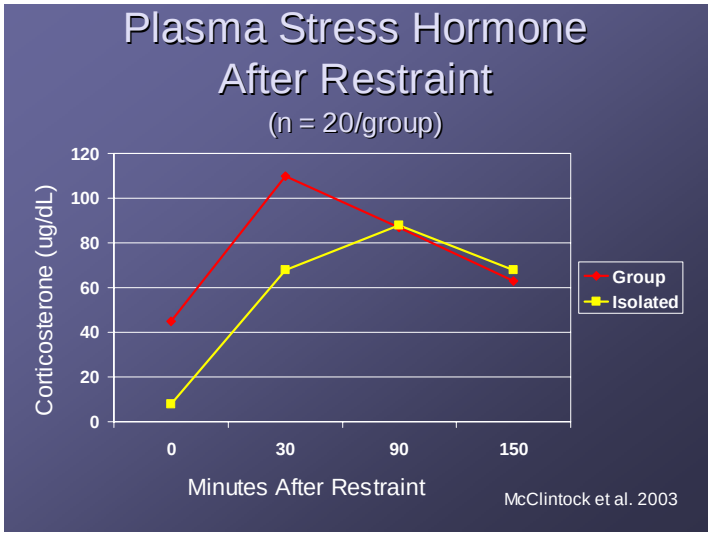
I believe that this new approach to urinary hormone analysis has potential to shed light on health disparities research related to heart disease mortality, breast cancer mortality, and chronic disease prevalence.



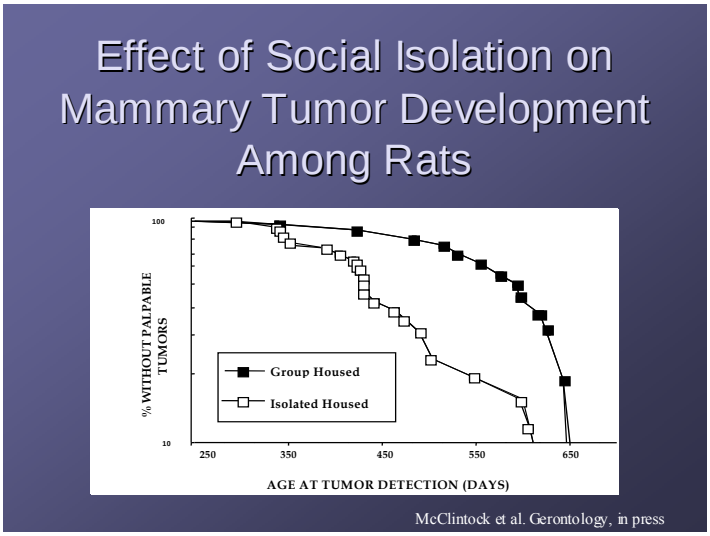


I'm sure you've all seen that heart disease mortality is higher for males compared to females and higher for blacks compared to whites in all age groups, and obviously there are multiple factors which contribute to disparities in heart disease mortality. Our results suggest that men may produce higher levels of epinephrine and norepinephrine compared to women. Do these differences contribute to gender differences in heart disease mortality? Our results also suggest that larger samples sizes may reveal significant racial and ethnic differences in epinephrine, norepinephrine, and cortisol production. Could such differences also contribute to racial and ethnic differences in heart disease mortality?

These are results from a recent study showing breast cancer mortality is higher among black women compared to white women. This gap has actually increased over the past decade. This is the topic that Martha [McClintock] and I and several others are pursuing through the University of Chicago's Center for Interdisciplinary Health Disparities Research, which was established a few months ago.



Martha recently found that rats raised in isolation have lower basal cortisol compared to rats raised as a group. In response to stress, both isolated and group-housed rats have a vigorous response, with cortisol going up, but in socially-isolated rats, we see that the cortisol release persists for a longer time period. So the issue of cortisol release, basal cortisol, patterns of cortisol release, is a fascinating topic. I believe studies that measure cortisol several times during the day will shed light on group differences in cortisol release. Because overnight urine samples provide an integration of hormone production over the previous 12 hours, assays which correct for residualized creatinine values may also offer valuable opportunities to evaluate group differences in hormone production.



Martha has also shown that isolated rats develop breast cancer at a higher rate than group-housed rats.

Disease Prevalence

	Female	Male	Ratio	Black	White	Ratio
Fibromyalgia	3.7%	1.0%	3.7			
Chronic Fatigue Syndrome	2.48%	1.93%	.8-1.33	3.38%	2.39%	1.28
Rheumatoid Arthritis	2-2.7%	1.5-1.8%	1.5-2.5	1.9-2.5%	2.1-2.4%	~1
Asthma	9.1%	5.1%	1.78	8.5%	7.1%	1.19

As far as other chronic diseases, basal hypocortisolism has been associated with several chronic diseases, including Fibromyalgia, Chronic Fatigue Syndrome, Rheumatoid Arthritis, and Asthma. We see that these are more prevalent among females compared to males. We found evidence of lower cortisol production among females compared to males, and so a question arises as to whether this difference in cortisol production influences the incidence and prevalence of these diseases. As you can see, there is some evidence of ethnic differences in the prevalence of these diseases as well.

Conclusions

- Creatinine correction of urinary hormones underestimates values among those with higher muscle mass and Blacks
- Residualized hormone values reflect correction for FFM and urine concentration
- Males excrete more epinephrine, norepinephrine, and cortisol compared to females after accounting for FFM and urine concentration

- There is some evidence of higher epinephrine and norepinephrine and lower cortisol among Blacks compared to Whites
- Gender and ethnic differences in catecholamine and glucocorticoid production may contribute to gender and ethnic disparities in heart disease, cancer, and other chronic diseases

In summary, 24 hour urine collection is problematic because it is difficult to collect and compliance is low. Overnight urine collection is also problematic, or has been in the past, due to this issue of variability in creatinine excretion due to differences in muscle mass. We hope that this new approach of accounting for creatinine production will allow us and others to move forward with analysis of overnight urine samples.

Lindau: For the NSHAP study we've eliminated doing stress hormones because of concerns about validity. We thought 'we can't do it.' Would you do it with three thousand people? Do you have the confidence in the kinds of corrections that you showed to really make use of those data?

Masi: The collection of 24 hour urine samples for hormone analysis is still the gold standard. Our study is really a pilot study. A follow-up study would collect 24 hour urine samples which would be categorized by time of day collected. Analysis could then be conducted on both the overnight samples and the 24 hour samples. This would permit a determination of whether the approach we propose for overnight urine samples yields results which are similar to results from 24 hour urine samples. So I think it's a nice pilot, but additional work needs to be done before the approach we propose is tried on a national sample.

Minimally Invasive Specimen Collection for the Detection of Drugs and Alcohol

Speaker: Christine Moore

I'm going to talk about minimally invasive specimen collection for the detection of drugs and alcohol. The byline to my job, actually, is 'blood, sweat and tears.'

What they told me was to share some information on new developments, to discuss some challenges for collecting certain types of biomarkers in population-based research, and to suggest a couple of future directions where the biomarker collection technology may be going. My laboratory started off as a meconium testing laboratory, which is a new-born stool, it's like a once-in-a-lifetime sample that tells you about drug exposure on the fetus from the mother. But from that, doing unusual samples, we've branched off into some other things.

What I'm going to talk about today mostly is hair, oral fluid, and of course there are other things that I'm not going to spend a lot of time on, because I know we don't have a lot of time. But you can test sweat, fingernails, toenails, you can test anything that comes out of your body. And people do. Earwax, and really, anything at all. And especially if you're dead, you can take any sample you want to test for drugs and alcohol. So what will these different things tell you?

Well, the difference with the samples is the time that you can cover with a sample collection. Everybody pretty much collects blood and urine. Now, blood's gone down quite a bit since the HIV issue, there's an infection risk, so most population people like to collect urine, they're comfortable with that, and I'll explain that in a minute. But maybe two or three day's worth of drug history is all that that will give you. Oral fluid, about 12 hours for marijuana, one or two days for other drugs. And as you go up, you can see that hair gives you about 90 days worth of drug history, about three months.

The USA is about 6% of the world's population, and about 60% of the world's drug market. I thought that was kind of interesting. About 20 million people, mostly in California, I think, admit to using marijuana -- admit to using marijuana -- so there's plenty that don't admit to it. About 6 million cocaine, half a million heroine, and about 13 million registered alcoholics. So that's not just people that drink, that's alcoholics. And I got that data from Drug-Free workplace website. We use it in our laboratory.

Why do we use drug and alcohol testing? Well, mostly poisonings, alcohol especially is a big factor in domestic violence investigation and auto accidents, and parole violations -- that's the sweat patch that if I have time, I'll get back to. You can use them in life insurance. If you need life insurance, they'll take a sample from you, most likely saliva these days, but used to be urine and blood. They will test you for cocaine, cotinine, for smoking, and alcohol markers to see if you are an alcoholic. And then they'll give you your premium, depending on the results.

We use drug testing, as I said, for screening of pregnant women, diagnosing fetal alcohol exposure, fetal alcohol effect. It has also been used a lot more in medical professional licensing -- people that want to get their licenses back, having lost it through alcohol or drug abuse. So a lot of testing goes on with that, and some determination of relapse.

Everybody uses urine. Why? It's the most widely used and accepted. Everybody knows all about it. You get a lot of sample, so if something happens you can do it again. There are lots of established collection and analytical procedures. The government, the state, all kinds of people will sell you proficiency samples, to the laboratories, so you know what you're doing, they know what you're doing, and they'll all come out and inspect you.

But urine doesn't really give you any information on whether the person is under the influence of the drug, when they took the drug, how much they took, none of that. It just tells you that there's drug in their system. It's incredibly easy to adulterate. If you go for a job interview, they give you a urine test, and you fail, you don't deserve the job. It is so easy to beat a urine test. And diluted, substituted, and invalid samples all cause big problems in the transportation industry. That has to do with creatinine issues as was being discussed earlier. The collection is not observed, and that's why it's so easy to beat – they don't watch you. And you can't really determine how much drug was taken.

Suppose we don't want to do that, we want to do hair. Well, this fits right in with today: simple, rapid, minimally invasive – depending on how much hair you collect. It will give you a long window of exposure, about 90 days, depending on the length of the hair. Of course, if the hair's enormously long, you can go back more time than that. But most laboratories will cut this off at 90 days. It is very easy to ship and to store. No known infection risk. And just in comparison, hair will give you a much longer history of drug use. Drugs get into the hair from blood, from sweat and sebaceous glands, and from the environment, which, of course, leads mainly to the first disadvantage of hair.

Because you get drugs in the hair from the environment, you can't be absolutely certain that this person used the drug. They may have been around somebody else using the drug. This is, of course, you would think, a huge disadvantage. Well, it is kind of a disadvantage, but it's a big advantage in some of the things that we do. We do a lot of child protection work, and they will send samples from children's hair that live in houses where these methamphetamine labs are, and cocaine. So it's great that we can find that they've just been around it, they didn't actually ingest it.

There's definitely an inherent color bias to hair testing; for basic drugs like cocaine and methamphetamines, not marijuana, they will incorporate much more into darker hair than lighter hair. That translates into a racial bias – it's not actually a racial bias, it's just unfortunate that African-Americans and Hispanics have darker hair than do Caucasians, for the most part. And Hispanic Caucasians have dark hair, too, and it's a color bias. The drug concentrations are very low, especially marijuana. It's very hard to find in hair. Marilyn Huestis from the National Institute on Drug Abuse gave a talk about two weeks ago here in Chicago, and she said that daily smokers of marijuana are not detected at the current cut-offs. So if you're going to do marijuana, don't pick hair. Confirmations require high sophistication, it's more expensive than urine or oral fluid, and you can adulterate it. I'll talk a little about bleaching in a moment. And the 'follically challenged' -- if people are bald, then you're not going to get that hair, because then it's not minimally invasive. You have to find it elsewhere on the body, if it all.

So as I mentioned, drug powders and smoke can be incorporated into the hair from the environment. Solvent washing certainly can remove some of the obvious powders, residue, mousse, gel, that kind of thing, but the drugs do go into the hair through sweat.

A few suggested solutions to this. There are washing procedures that you'll probably find in the literature. The problem with that is, not every lab can do them, in fact no labs can do them, there's only one lab that says they can wash off all the outside drug, nobody else can do that, so it's not a reproducible effort. It's a good idea, but it doesn't seem to work. This is very arbitrary. Above such a

number would be positive, and below such a number would be negative. Above, you've used; below, you didn't. It's okay, but it's not satisfactory. This is the best way to go, to find some metabolites in the hair, from the drug, that only got there by you taking it, physically inside, so you have to look at metabolic profiles for all of these drugs, and not just cocaine.

Some work from Utah, Salt Lake City, the Center for Human Toxicology, they do some great work up there. They have rats with different color hair, like black and white on the same rat. So they give it drugs, and then measure the white bit and the dark bit, and they also transplant that hair from one rat to another, it's amazing. But they've shown in animals, in vitro and in vivo, that darker hair incorporates more basic drug than lighter hair, and that's because of the nitrogen. This bias is not an issue with acidic drugs, so for barbiturates, marijuana it's not an issue.

Bleaching, perms, and dying your hair do cause drugs already in the hair to degrade. But if you bleach your hair before you take drug, you've made it more porous, you've made it wider, so you're going to get more drug in, so that's a little catch-22. But more drug will be washed out with normal hygiene.

So hair is being continually accepted. The SAMHSA – Substance Abuse and Mental Health Service Administration – is currently working towards allowing hair, oral fluid, and sweat to be included in the federal workplace program. If you have a federal job right now, the only specimen they can legally take from you is urine. Hair is non-invasive, easy to ship, gives you a bunch of information over a long period of time, and is legally defensible, when confirmed, and under consideration by the federal government.

Saliva – well, oral fluid – is simple, rapid, and non-invasive. It will give you evidence of being under the influence, so it's great for roadside testing. It's great for a man who crashes his truck at work. Do a saliva test -- what's the use of doing a hair, or urine? That isn't going to tell you that he was under the influence of marijuana when he crashed his truck. Do a saliva or a blood, but a saliva's much easier. It has a good correlation with blood-alcohol, and it's hard to adulterate. You really don't want to walk around with something other than your own saliva in your mouth. It's really hard to do that, so it's hard to adulterate any kind of sample. And the insurance industry is really getting into this. Nowadays, for a life insurance sample, they'll come into your house and they'll take a saliva sample. They don't bother with the blood anymore. You really don't need any special collection facilities, you don't need a medical person. Here are more things for you to see – saliva collection devices.

There are low concentrations of drug, there is a limited or unknown sample size, and it's a much shorter window of exposure than urine. It will only give you recent use, and marijuana is a problem for this. And, again, it's more expensive.

The problem with saliva is how you collect it. There are a couple of major ones on the market, the Orasure Intercept and the Quantisal. Effectively, the problem with that is, you really don't know how much oral fluid goes into these. I've passed it around, and as you'll see, their own literature shows 50 to 800 microliters collected. That's a pretty big range. Was it 50 or was it 800? They don't know. It could be anything in there. So you really don't know how much oral fluid you got, so your cut off concentrations don't mean anything. The Quantisal is marginally better. They have an indictor for when you've collected 1 millileter, but some of the drugs really just stick on that pad, which we then put into the buffer. And you can spit in the cup, but not everybody likes that.

So there is a definite sample source and observed collection. And it is minimally invasive. It is the preferred biological sample for most people. There is a 90% cooperation rate for oral-fluid in population-based studies compared to 76 for urine and 69 for hair. And that's Dr. Fendrich's article in

Addiction in 2004. So if you want to do a big population-based study and you want a lot of participation, saliva might be the thing to do. There are minimal collection skills required, and limited chance of sample adulteration.

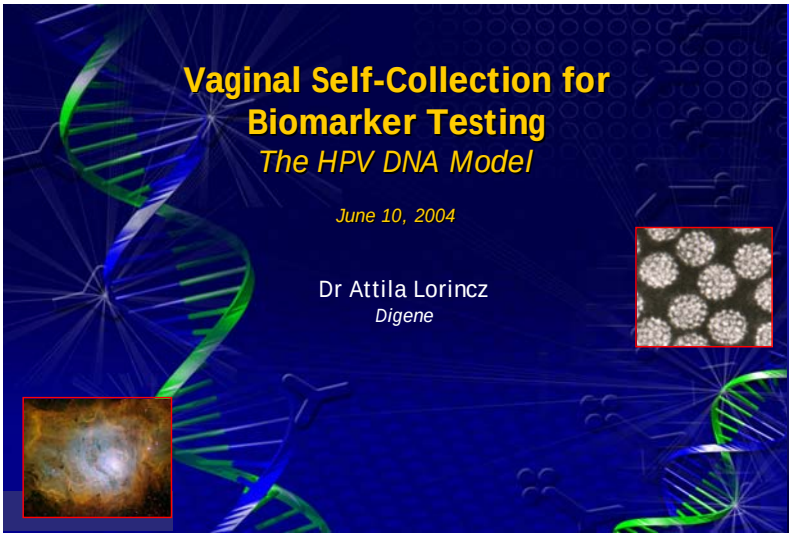
And that's basically a comparison of the specimens, as I've mentioned. Sweat, briefly, has one application: it's an elastoplast type of sweat patch that goes on your arm or your chest or your back, stays there for a week. You can shower with it on, you can run with it on, you can do whatever you want. If you take it off, you can't put it back on, so that's how they know if you had it on all week. It gives you about seven days of history, because it's a week, and it's for parole violation. That's the only good application there is for this thing. Nobody else is going to wear a patch for a week, and then come back. Drug abuse is a parole violation, so they'll put the patch on, come back in a week, and they'll test you. There's obviously no time to go into the rest of it, so I'll finish up there.

Lindau: I have a question regarding alcohol. You mentioned that there was a test for pre-disposition to alcohol use?

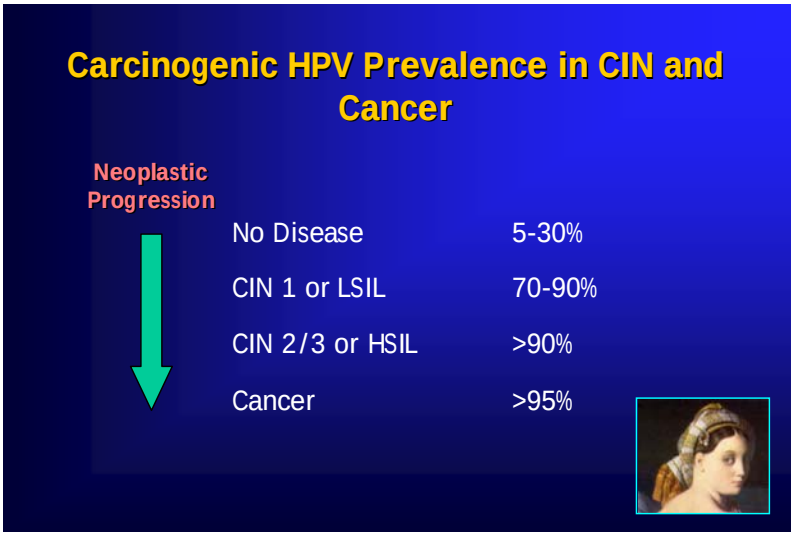
Moore: Oh, that's genetic testing. We don't do that testing, but there are genetic tests for predisposition.

Vaginal Self-Collection for Biomarker Testing: The HPV DNA Model

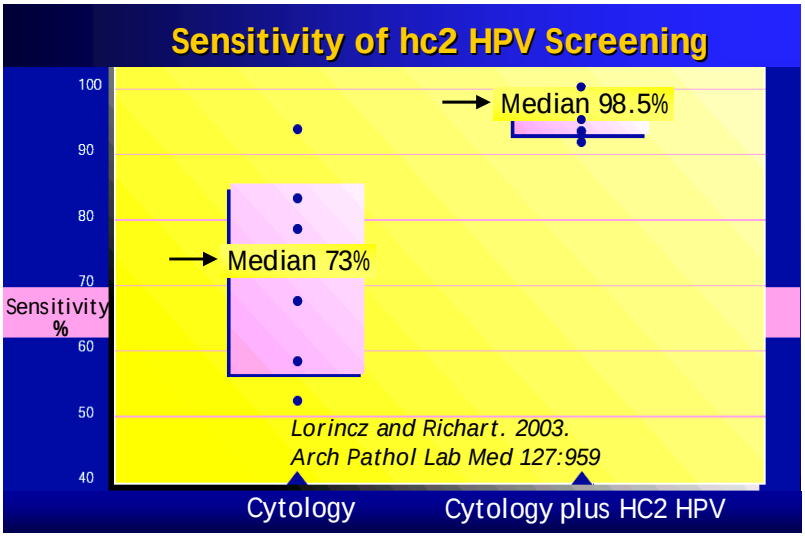
Speaker: Attila Lorincz



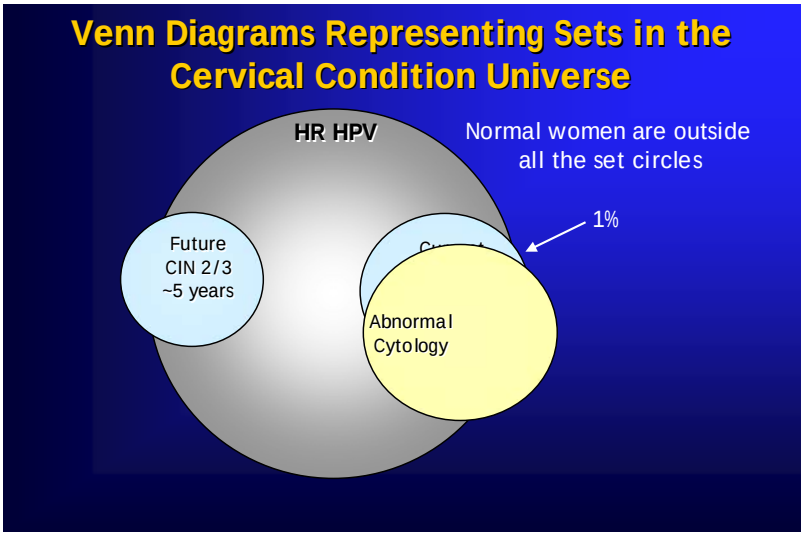
I'm going to talk mainly about HPV and self-collection from the vagina. Although I have some experience with Chlamydia, Gonorrhea, and Herpes, the issues are somewhat different depending on what you're trying to determine, and I'm going to point those out as I go along.



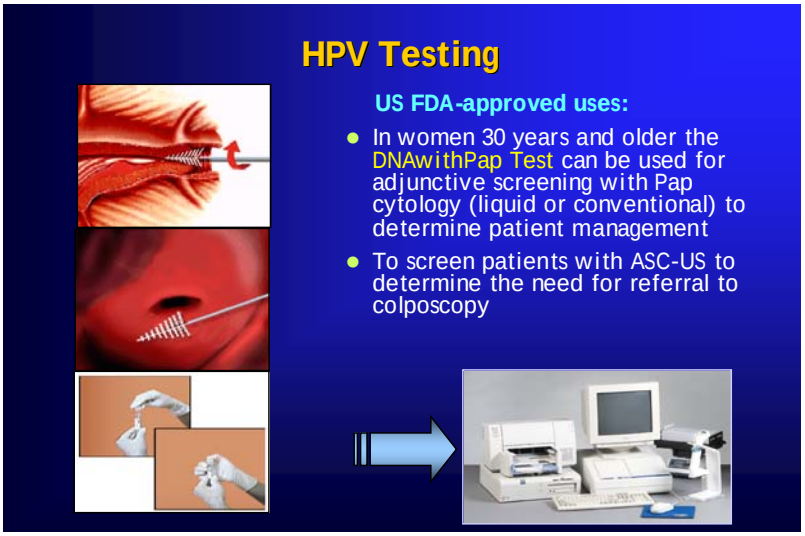
I have a few comments to make on HPV testing, which I think will put it in perspective. HPV, of course, is a group of viruses that are involved in the development of cervical cancer, and this slide summarizes the kind of data that has come out of hundreds of papers. Most of these studies were done using cervical sampling with the speculum assisted pelvic examination. Typically what you will see, depending on the ages of the population, is a rate of HPV positivity anywhere from about 5 to 30 percent, or maybe higher, in younger women. However, in women in the age groups 40, 50, 60, HPV positivity drops fairly low to about 5 percent at the cervix. It's a little bit higher in the vagina, interestingly. Then as you get to higher-grade disease, such as CIN 2/3, cervical intraepithelial neoplasia grade 2/3 -- or cancer -- the positivity rate for HPV, carcinogenic types (now bear in mind, we're talking about a group of 13 types here, in fact there's probably something like 17 or 18 that may need to be in what I would call the optimal cocktail), positivity goes to about 95% to close to 100%.



Here's some data that I reviewed with Dr. Richart that shows you, for example, the performance of cytology. Cytology is the typical methodology for detecting high-grade disease, that includes CIN 2/3 cancer. The median sensitivity there for detecting those pre-malignant lesions in cancers is 73%. A combination testing of HPV plus cytology gave a median value of 98.5%.



This brings together, in a simplistic schematic sense, our, at least my, current understanding of HPV infection. If we think of the HPV infected universe as a subset of all women, let's say 20%, 30%, or perhaps 5% prevalence, depending on the population, a subset of the women have current, high-grade disease. A vast majority of these lesions are HPV positive. Only a small sub-segment are negative. Abnormal cytology picks up some of that, but leaves undetected a large segment that is detected by HPV. If one combines the two methodologies, there's only about 1% that are left that are not detected by either cytology or HPV, and there is another large group of women who are going to develop the disease within 5 to 10 years who are not detectable by cytology. I draw them over here just for representation. They evolve out of the HPV positive group.



Now, all or the majority of this data has been determined using a speculum-based cervical examination, where a brush-shaped device, or some kind of a broom, goes into the cervical os, guided by the clinician or the nurse, and is rotated, placed into a tube, and it is sent off to the lab. This particular little brush-shaped device can also be used for self-sampling. Brushes can be used, and various other kinds of devices.

I'm going to talk about self-collection issues. The types of devices that can be used are either swabs, brushes, tampons, or other kinds of devices. How about the issue of collecting from men? Well, that is somewhat of a problem, obviously. I would suggest a possibility, and I don't know the acceptability of this, but anal collection in men may be appropriate under certain circumstances. Probably not for all men, but certainly for gay men. There's a very high rate of anal cancer that develops in gay men, and also in women, so that is an alternative site. When we talk about the best devices here, some of these can be problematic. Tampons can be used, but it's very hard to get material off tampons. Swabs, by and large, are the most acceptable. Dacron swabs are quite good, cotton swabs are not because the fibers come loose. Brushes yield a very good sample, such as that small brush device that I demonstrated. But women find brushes slightly less acceptable than swabs. There are other kinds of devices, which are hybrids of either swabs or brushes, with an outer sheath, so that the device can be inserted into the vagina and then it can be pushed through in an attempt to reach the cervix. Some of these are fairly exploratory at the current time. And, of course, the issue of cost comes into play here, especially for places like India, China, the developing world, where self-sampling, I think, is a very interesting modality. A brush or a swab is a very simple, inexpensive device. Some of these things can be fairly tricky in terms of cost and manufacturability.

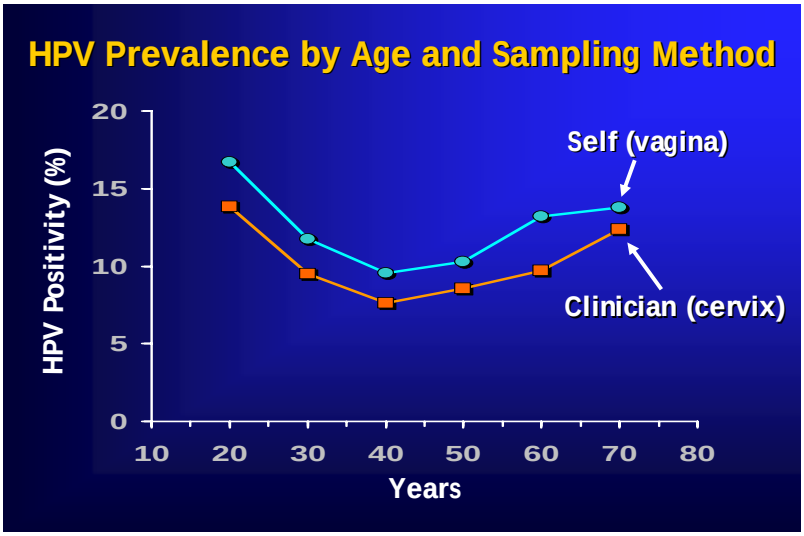
Purpose. The purpose of a self-sampling, vaginal self-sampling, let's say, would be to collect cells and fluid, and from that you can get DNA, RNA, protein. And you can preserve the morphology as well, depending on the particular sampling medium that you take. The procedure is relatively simple and easy to teach, and I'm going to get into that a little bit.

Acceptability. There are issues about utility, utility for the lab, and the researcher or the clinician. Safety. Can this cause patients damage? Adverse events? It's a fairly interesting topic. The Morelos study in Mexico, which I'm going to talk about, is a study of 8,000 women, and the majority of women found the device to be acceptable. There were less than 10 adverse events, several of them involved a device getting lodged in the vagina, where the nurse had to take it out. There was one case where an elderly lady took the sampling device, took her self-sample, and then drank the transport liquid, for whatever reason I can't imagine, and then gave the nurse the swab. So she obviously had to go to the hospital and have her stomach pumped. Not that the medium is terribly dangerous, but it is somewhat irritating. So those are the kinds of things that could happen. Dry devices are an interesting possibility, where the danger of the patient drinking, or spilling it on themselves, or pouring on their eyes or some crazy thing that they might think they have to do, is really minimized.

Then we have the issue of costs and stability. Stability depends on DNA, RNA protein morphology. And of course, data. There are a number of studies that have been done. There's study data for HPV. I have some information on Chlamydia, gonorrhea, and a little on herpes toward the end.

This is a simplified self-collection instruction that we've actually used in our studies. As you can see, these are hand-drawn. More sophisticated versions of these are available on posters and so forth. We actually used this in the study in Shanxi, China, the SPOCCS study by Belinson et al. And of course, there was wording in the local language to go along with these. It shows you the device, a woman taking the brush, she inserts it into the vagina in a standing sort of semi-crouched position, pushes it all of the way up, attempts to reach the cervix, but that's not really terribly important, and then places it back into the tube.

Here's some data from the Morelos study. It was conducted in the state of Morelos, close to Mexico City. Here is the HPV prevalence by age and sampling method. Now this just looks at positivity for HPV using the 13 HPV types of interest that are in the hybrid capture device. As you can see, there is an interesting age trend, although in this population, it tends to dip up in the elderly women. We think that there's some sort of an immune-suppression going on in these women, so that a lot of HPV that they get early on becomes re-activated. There was a higher HPV positivity in the women who sampled from the vagina.



Now, when you look at the correlation between these two, 50% of the positives were common between cervix and vagina, and the other 50% were discrepant. Of the 50% that were discrepant, approximately 2/3 were positive in the vagina and not at the cervix, and 1/3 was positive at the cervix, and not at the vagina. So if there is variation in the anatomical location of HPV infections for example, perhaps that might also be true for other infectious organisms, in that depending on which areas are infected might lead to sampling and test variability. Variability may well be less so for many other kinds of biomarkers. If you're interested in DNA, of course, you'll get very good DNA from there and it will be representative, unless you're looking for somatic mutations, which might be more regionalized. If you're looking for RNA expression, then depending on tissue location you might expect more variability. So it's not a totally clean situation.

Morelos Study Test Performance Comparisons

Detection of CIN 2/3 (99) or Cancer (8)

Test	Sensitivity%(CI) n=101	Referred%(CI) n=7,736
Pap smear	59 (49-69)	2 (2-3)
HPV SS	71 (61-79)	12 (11-12)
HPV CS	93 (86-97)	9 (9-10)

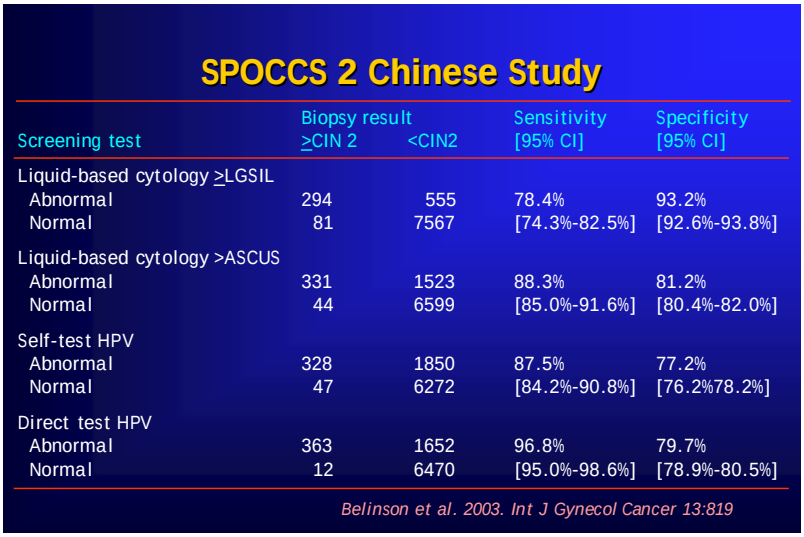
If we look at the performance now, not just for detecting HPV but an ability to detect high-grade CIN, or cancer, the story is different because the cervix is the location of 90% of the cancers and high-grade lesions we're interested in. And in fact, the cervical sampling technology detected 93%. In this study, self-sampling at the vagina detected 71%, and the Pap smear in this case detected 59% of the approximately 107 high-grade lesions and cancers. So you can see some variability. There are some other self-sampling studies that have found higher values. But I can tell you that in general, it's my impression that self-sampling will have about a 10% lower sampling sensitivity. The other point is that it actually has a somewhat higher rate of HPV positivity, which I pointed out in the previous slide.

Women's Preferences

Morelos Study

In a nested study of screening procedure acceptability among 1,100 women, 65% reported that they strongly preferred the self-sample for HPV over the pelvic exam and Pap smear.

What about the preference of women? Well, in the nested study here 65% of the women reported that they strongly prefer self-sampling HPV over the pelvic exam for HPV. When you ask the question a different way, what percentage of the women will be willing to accept self-sampling, then greater than 90% said 'sure, it's acceptable.' Now, in other locations, in Uganda, in China in the SPOCCS study, about 90 to 95% of the women actually found it acceptable and provided a sample. In some populations, perhaps based on religion, there may be difficulties. We're a little bit concerned about certain Muslim groups or certain macho societies where men might get involved or there might be some sort of taboo against self-sampling. And strangely in some questionnaires that we looked at, some younger women, especially undereducated or lower socioeconomic level, felt that self-sampling wasn't as acceptable -- they didn't want to touch themselves there. Or perhaps they felt that a clinician could get a better sample; they were concerned about not getting a good sample.



In this other study from China, as you can see, the self-sampling sensitivity for high-grade disease was quite a lot better than in the Morelos study, detecting 87.5% of the high-grade lesions compared to a 96% sensitivity for clinician-assisted sampling. And the liquid-based cytology, read by an expert, was either 78% or 88% if the cutoff was greater than or equal to atypical squamous cells of undetermined significance.

Finally, issues of cost, stability, and so forth. The cost of a swab-based device is probably on the order of about 50 to 75 cents. If you just have the tube and the swab, and you omit the liquid, I guess that could be 30 to 50 cents at the lower end of the range. If you're using brushes or some of these more sophisticated devices, the sad fact is it's probably going to be higher, especially if you're using a sheath-type device, and may be over a dollar.

In terms of stability, these samples are stable for very, very long periods of time, especially when you talk about DNA, RNA, and linear epitopes in proteins -- several months, up to nine months. Morphology is quite stable. You can do immunocytochemistry off these. So I think they're a very good device. For Chlamydia and gonorrhea sampling, urine sampling has been the predominant modality, but there are a lot of papers out there now talking about vaginal self-sampling. The studies find a fairly high level of equivalence to cervical sampling for Chlamydia, gonorrhea. So I imagine vaginal self-sampling will be a good way of getting a sample for STDs, for other biomarkers, for RNA expression. And I'm pretty bullish on this. I think it can be very useful for global healthcare.

McQuillan: NHANES is using this and we find a very high response rate after we hit age 16. The 14 to 16 year olds are very reluctant, especially if they don't use tampons. But older women, when we do the 50-59 year olds, it's like 98%, because they really buy into its importance, so we're getting good results.

Lindau: Do you see any possibility for using serologic antibody testing for HPV exposure in men and women?

Lorincz: Yes, there is some value for research purposes. I would say the preferred medium there is a serum-based test, although you can find transudated antibodies in the vagina, and in self-samplings. And in fact it correlates extremely well with what's in the serum. Philip Castle did a nice study. The issue with serum-based sampling for HPV is that it actually correlates fairly poorly to true disease. There's a very long immunological memory in people who've been transiently exposed. So you get a very high seropositivity in women and men who've been sexually active, regardless of whether they actually have the disease or not. By the same token, some people are non-responders, and about 25% of people do not become seropositive even though they have an active lower genital tract infection demonstrated by PCR or hybrid capture. So I would use it with caution and sparingly, but it can be quite interesting. You can also determine type-specific responses because there are specific serological panels available for telling if it's 16, 18, or whatever.

McQuillan: We're doing that with NHANES, so we might get some information for you.

Estrogen, Sexuality and Health

Speaker: Martha McClintock

Much of our discussion of biomarkers has focused on non-invasive measurement of hormone levels in blood. But hormones do not exert their effects in blood, they must be bound in cells of specific tissues. The argument I want to make is simple: let's go right to where the action is. Let's look at the cells themselves. We are proposing to measure the vaginal epithelial cells not only as tissue that is directly involved in sexuality but also as a biomarker for other aspects of estrogenic activity.

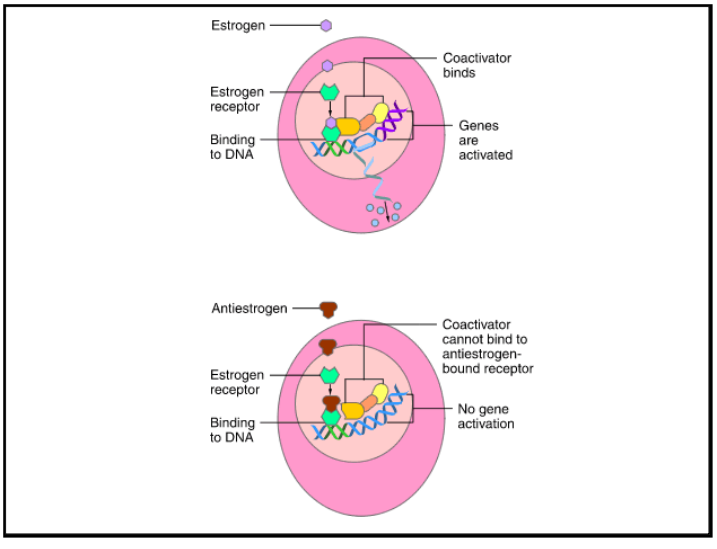
Estrogen increases
cytokines and immunoglobulin
in vaginal compartment

IgG	I-1 beta
IgA	IL-6
	IL-10

Al-Harthi et al, J. Interferon Cytokine Res 2000 20:719-24
AIDS Res Hum Retroviruses 1998; 14:S51-5

Estrogens have a direct effect on vaginal elasticity and lubrication during sexual arousal as well as risk for vaginal infections, by regulating cytokines and immunoglobulins within the vagina. In turn, vaginal condition has a very profound effect on women's sexual practices and desires. Estrogens also affect the brain, modulating not only sexual desire, but a variety of health related functions such as sleep quality, memory, immune function, and processing of sensory stimuli. So, not only are we interested in what the actual condition of the vagina is, but using it as a window into what the bioactivity of a whole suite of hormones is.

Estrogens, estradiol 17-beta being the most potent, exert their effects by passing through the lipid cell and nuclear membranes and binding at an estrogen receptor. This complex, the chartreuse and the purple, then bind to the DNA, regulating promoters and production of mRNA. The estrogens in the blood can not do anything without binding to receptors within the cell! Thus, antiestrogens used in cancer treatments act by blocking the estrogen receptor, thereby preventing activation of genes, i.e. those involved in mammary cancers.

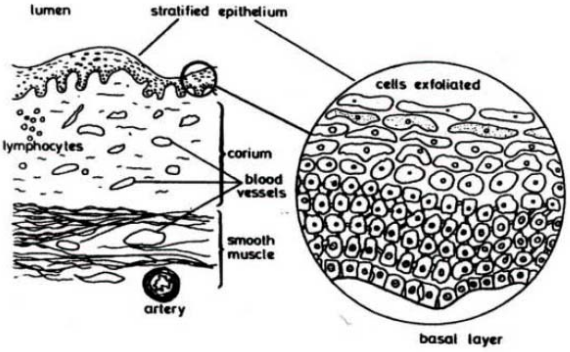


Androgen Insensitivity

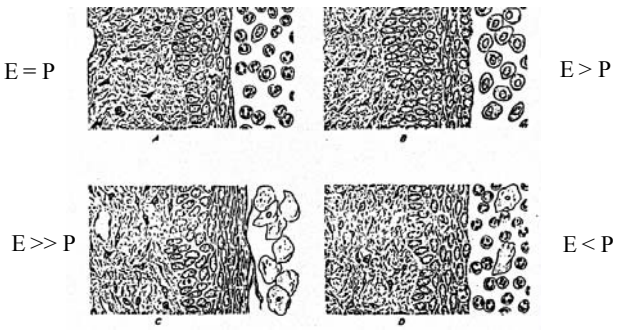
To bring this point home vividly, here is a woman. She has androgen insensitivity syndrome. These are her gonads, her right testes, and her left testes. She has testosterone levels higher than a teenage boy. But until recently she could not participate as a woman in the Olympics, because she's XY genetically. However, morphologically, in terms of her muscles and secondary sex characteristics, none of her cells can see the androgens that her testes are producing, because she has a genetic mutation in the androgen receptor. Her tissues do see the small amounts of estrogen converted from testosterone, which is just one metabolic step away, and so she's highly feminized and obviously a woman.

Obviously we can not measure the presence of receptors on cells in survey research. So how are we going to do it? Look directly at the vaginal tissue, which is exquisitely sensitive to estrogen

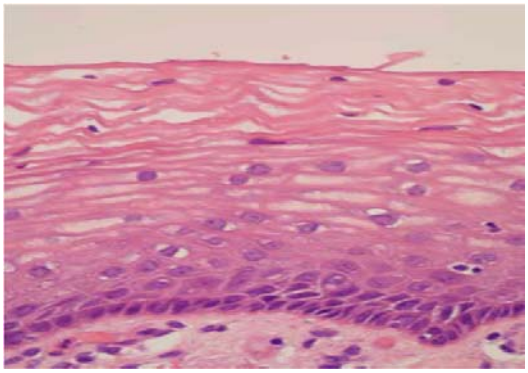
Here is a cross-section of a vaginal tissue. Here's an artery, the smooth muscle, then there're blood vessels in here, with lymphocytes. These are immune cells, this is the lumen of the vagina up here, with the stratified epithelium, which is enlarged in the insert. You can see is that as you go up layers, the cells change in shape, and become flattened and then actually become exfoliated. It's the exfoliation of these cells that's the major source of vaginal lubrication, in a non-sexually aroused woman.



Original Bioassay for Estrogenization

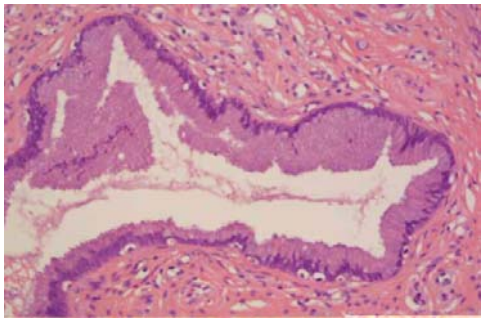


Looking at the cytology of vaginal tissue was the original bioassay of estrogenization, before radioimmunoassay was invented. This is what we measure non-invasively daily in our rat colony to track the changes in ovarian cycle. Progesterone also affects the vaginal epithelium, and it's the relative balance of these hormones that determine the functional level of estrogenization. Progesterone regulates the number of estrogen receptors. But, for the purpose of survey research and assessing the estrogenization of the vaginal epithelium, we need only look at the cells that slough off naturally, and from that we can get a very good idea of the level of functional estrogenization.



Vaginal Wall Cytology

Here's cytology, for those of you that want to see what the real cells look like, that those cartoons are not just artists' conceptions. It's really beautiful to look at under the microscope and see the layers.



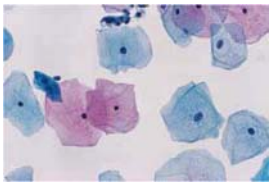
Mucus Secreting Gland in Cervix

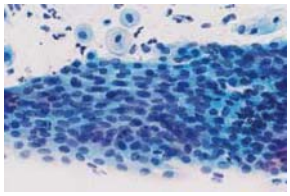
Here is a cross-section of a gland that secretes mucous, which is another source of fluid and lubrication in the vagina.

What's really exciting is that, in the NSHAP study, we're going to be taking this well-validated classic method of estrogenization out in the field, and determine how it correlates with important psychosocial measures associated with sexuality and health.

Here is a premenopausal, estrogen-dominant sample of cells as they appear through the microscope. They look very different from postmenopausal cells, which are mostly parabasal. These cells are at the bottom of the vaginal wall, and are the ones to slough off when estrogen is not present. Without estrogen, the tissue is not thick and is not elastic, to the point of vaginal atrophy.

Premenopausal	Postmenopausal
Estrogen Dominant Low Progesterone	Low Estrogen Low Progesterone
Squamous cells	Parabasal cells

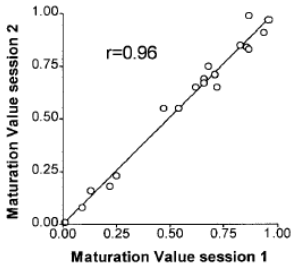




Dr. Lindau and I went around and around on the topic of validity and reliability of this method. 'What's the validity?' I replied 'This IS what matters biologically. It is the estrogen assays and methods for counting receptor number that need to be validated against this real biological phenomenon!'

What is the reliability? A common measure of estrogenization is the maturation index. Reliability is assessed by having different readers look at a set of slides in one session, score them and then compare the two readings. Here, there is a very high reliability between an automated cell recognition and a highly trained histologist.

Here is a study from 19 women that raises interesting issues. Notice the variation across women, that is, the individual variation in the sample of the Maturation Index (called Value here). Seven of these women were pre-menopausal, and ranged from 27 to 50 years of age. Then there are six post-menopausal women that have an atrophic vaginal epithelium, and they range in age from 55 to 75, they're down here. However, 6 post-menopausal women who've maintained a mature vaginal epithelium, ranging in this study from 52 to 68, who fall sort of in this range here. It would be very interesting to find out how the estrogenization of the vagina in these cases correlated with sexual activity, sexual desire, and sexual practices. Moreover, we can use this as a biomarker for estrogenization in association with other health states.



7 Premenopausal	27 to 50 years
6 Post menopausal Mature	52 to 68 years
6 Post menopausal Atrophic	55 to 74 years

Jeroen et al. 1999 Cytometry 35:196-202

Waite: What is the maturation value? What does that tell us? How do you interpret it?

McClintock: It is the proportion of different types of cells that reflect levels of estrogenization. The higher the value, the more estrogenized the vaginal epithelium is. One can not assume that higher is better, because, for example, it could be correlated with higher risk for breast cancer in addition to greater sexual activity.

Unidentified Speaker: It's an empirical question about whether it relates to anything about sex in older women. I think that's yet to be determined. One of my reasons for wondering that is that in the Women's Health Initiative, we have a grant, a clinical trial of estrogen and progestin, we certainly changed the maturation index but didn't change sexual function very much. So again, I think it's still is an open empirical question about its relationship to real human sex.

McClintock: Yes, it's an important question. I take issue with the conclusion of the Women's Health Initiative study that sexual function was not affected. The measures for sexual function were unidimensional. I think that this NSHAP study has an opportunity to really test that hypothesis in a powerful way.

We've done a study of changes in women's sexual desire over the menstrual cycle. There's no effect of the menstrual cycle on sexual activity. However,

there's a significant effect on sexual fantasies and desire, self-reported desire, and the Women's Health Initiative did not look at that. Here we will. It is also exciting because we'll have women presumably using different forms of HRT, and we'll be able to see how they are associated with a known, valid, reliable measure of estrogenization, as well as other aspects of women's health.

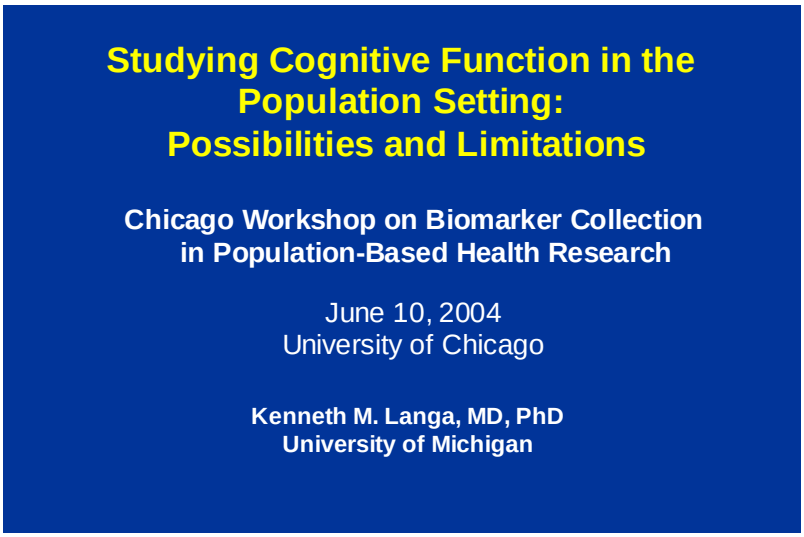
Rhodes: How did you collect the specimens for these 19 women, and how are you proposing to do it in this study?

McClintock: We are using the swabs discussed by the previous speaker and will be perfecting and validating a technique for getting a good suspension of cells to quantify.

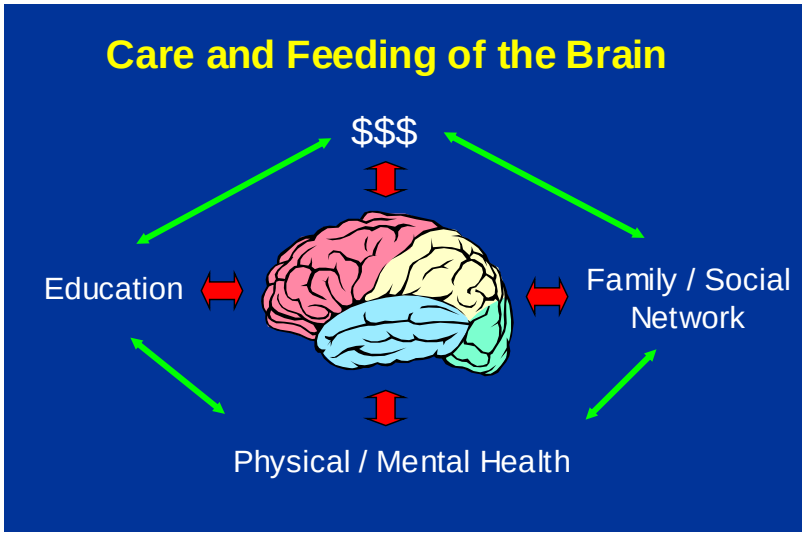
Studying Cognitive Function in the Population Setting: Possibilities and Limitations

Studying Cognitive Function in the Population Setting: Possibilities and Limitations

Speaker: Kenneth Langa



Langa: We're going to talk about care and feeding of the brain. I'm going to lay out some very simplistic ideas about how complex this issue is and try to sort through cognitive function and its effect on physical health, mental health, and social well-being.



I'm going to start with education. As you know, there is a wide range of research showing that education -- reading books, and thinking thoughts -- changes the actual biology of the brain, changes the number of connections and the stability of those connections as one ages. So education is very important in that perspective.

Education obviously also affects your human capital, as Gary Becker would tell us from Hyde Park. It changes how one can obtain and use information – an important issue in terms of the relationship between cognition and health. It also probably changes the likelihood that one will think about healthy behaviors, and the attractiveness of future benefits, so that is another important causal pathway.

And finally, it also probably affects your social network and whom you end up marrying and living with. So education is important from all those perspectives.

Then there's also money, or wealth, as we've talked about. There's an obvious connection and much research between social well-being and wealth, and social class and health – both physical and cognitive health. Again, perhaps through access to education, access to healthcare, and the neighborhood in which one lives, there is going to be a connection between your wealth and your brain's health. Also, in the other direction, again possibly through education as one of the pathways, your labor opportunities, your

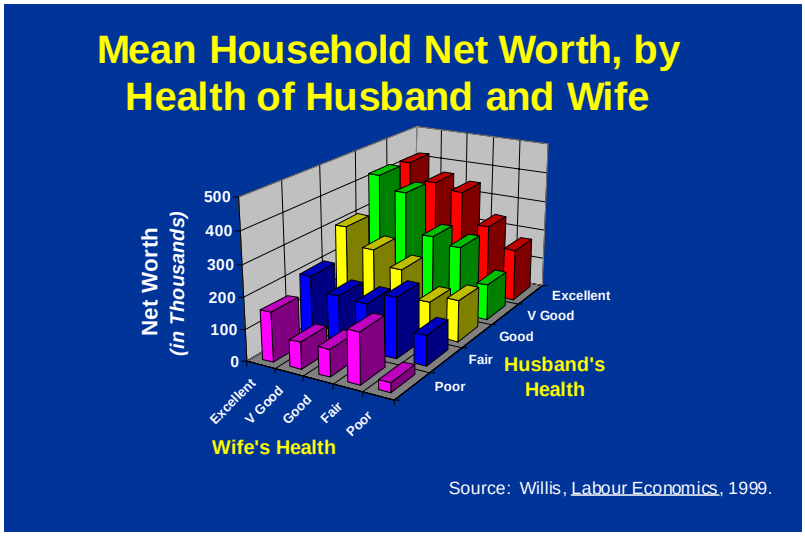
expected earnings throughout your life, will be affected going this way. So an educated brain probably will end up having more wealth. The person will have more wealth at their disposal. Some work that Bob Willis has done with the Health and Retirement Study, that educated brains, or brains that can think probabilistically, and think about the future using probabilities, might be able to gain more wealth over time. Maybe Bob will talk some more about that in his remarks.

Again, there are many different pathways here, and sorting this out is complex. The family and social network (as Linda Waite's work on marriage's effect on health and well-being has shown), social networks, the neighborhood that you're living in, and whom you're interacting with, will be important for maintaining health.

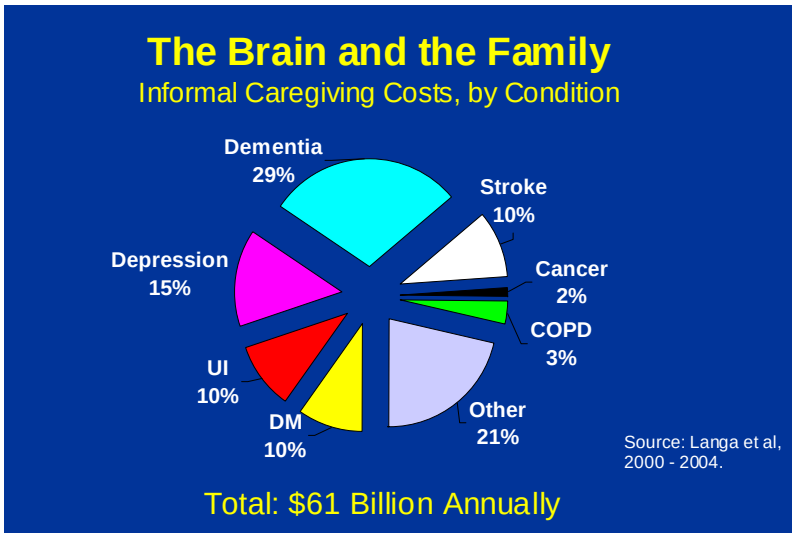
In the other causal direction, someone who is demented obviously needs more help from their family, which will lead to increased caregiving time, and perhaps caregiver stress which may negatively affect the spouse's well-being.

And finally, physical and mental health. Choose your pathway – how it gets that way – but perhaps through education again, the idea of investing in healthy behaviors, not smoking, exercising, will likely mean you're less likely to get vascular dementia, probably Alzheimer's disease also, as more research is showing.

The panelists have promised to totally elucidate all of these complex pathways for us by the end of the session. In eight minutes or less.



Two shameless plugs for the Health and Retirement Study: some work that Bob Willis has done just showing this health/wealth gradient from a couple's perspective. Along these axes here are self-reported health -- here's the wife's health, the husband's health. You can see there's a pretty clear gradient in the net worth, which is on this y-axis here, so the better the self-reported health, the more wealth you have, and it's sort of additive with the husband and wife.



And finally, some other work we've done on caregiving issues with the Health and Retirement Study. We've looked at the total amount of time that families spend providing care to people with various chronic conditions: urinary incontinence, diabetes, a whole bunch of other things here, lung disease, cancer, but you'll see the big three are all brain things. Dementia was the most significant one, about 30% of the care – what we've estimated as about $61 billion annually. Depression and stroke, too, so more than 50% of the care is due to brain-related issues.

Turning then to the questions that Stacy wanted us to address, I've laid out four questions here. First, 'What are the most important pathways between and among cognitive function, health, and social and economic well-being? What measures of cognitive function are most important to better define and study these pathways?

Questions for Discussion

- What are the most important pathways between and among: cognitive function, health, and social / economic well-being?
- What measures of cognitive function are most important to better define and study these pathways?

What are the significant or most important challenges to address when studying cognition in a population-based setting? And finally, Is collection of biomarkers for cognitive function in a population-based setting feasible and likely to help elucidate these pathways?

Questions for Discussion (cont'd)

- What are the most important challenges to address when studying cognition in a population-based setting?
- Is collection of biomarkers for cognitive function in a population-based setting feasible and likely to help elucidate cognition-health-wealth pathways?

Studying Cognitive Function in the Population Setting: Possibilities and Limitations

Speaker: Christopher Clark

Basically, with the issue of collecting information on cognitive function, how that can serve as a marker for the presence of dementing illnesses or the risk for developing a dementing illness, I've decided to try something new, so I have no Powerpoint slides, no visuals at all, with the hope that it'll keep it a little more interactive.

I want to do two things: I want to make about eight points, points that I think fall into the category of points of departure if you're going to try and think about these things, that is cognitive function as a marker for dementing illnesses. Points of departure that you have to be aware of, or at least have to deal with, if you're going to get interpretable and informative information. And then a couple of my own ideas on strategies that could be used in a community-based sampling of biomarkers that could be efficient and reliable.

First of all, points of departure. A lot of this stuff may seem obvious to you, but I need to get it out anyway. The first one is, of course, in a community sample (which is very different than a sample that comes to a clinic that is focused on the evaluation of dementing illnesses) the expression or the observation of cognitive impairment can be due to a large variety of reasons that we never have to deal with in a diagnostic clinic setting, and many of which have nothing to do with the development of dementing illnesses and operate completely independently of each other.

So the causes for cognitive impairment in the community are going to be much broader than the causes of cognitive impairment in a specialty clinic whose focus is the diagnosis of dementing illnesses. You need to be aware of that, and somehow filter out those non-dementing causes within the community.

I'll give you some examples. I think where it was presented best is actually in the Canadian Study on Aging, when they made cut-points on what they considered age and education-related cognitive impairment, and then made some attempt to define what those problems were. The very obvious one is mental retardation, or any birth-related injuries. If you fall below age and education-adjusted means for cognitive impairment, you're not going to be headed for a dementing illness, you'll have a static problem. It will be confounded by age, of course, because it'll become worse as you become elderly, but it's not a dementing illness. Medical problems and psychiatric problems, in the Canadian Health Study, schizophrenia was a significant contributor to performance below age and educational adjusted means for cognitive impairment. It's not a dementing illness. Other medical problems, including some brain problems like stroke – some of them are very obvious. But the point is you simply can't go out and sample cognition adjusted for age and education and say these people have a dementing illness, because you'll find a lot of things that aren't dementing. In a memory disorders clinic you can do that, because people don't come to you with static problems.

The second thing is somewhat related to helping sort that out, and that is that the public health issue for dementing illnesses is identifying cognitive impairment that represents a change. That is the single most important thing, and of course if you're doing a one-time community sample, defining change is not trivial because you only have your one shot at it and everything else is a historical issue. But that's the

single most sensitive measure that you have. It's not the cognitive performance of the subject, it's identifying who's cognitive performance has changed or declined.

The third point of departure really is related. We already touched on it, and that is that there are tons of medical problems that will produce cognitive change over time that are chronic: AIDS, low chronicity but still lethal; brain cancer or any CNS infections; and some psychiatric problems such as chronic schizophrenia. So change itself is not going to be adequate, as sensitive as it may be.

The fourth thing is that reliable information about the presence or absence, either one, of the symptoms of cognitive impairment cannot come from the individual, the subject themselves. You do not get reliable information. You get reliable information about cognitive performance from the individual, because they're actually performing, but the symptoms of cognitive impairment, confused thinking, judgment problems, etc., will only come from an observer. And that's what confounds the problem of community evaluation for presence or absence of dementia, dementia-related cognitive impairment. Because just as in the clinic, in the community you need a dyad in order to make that distinction. The only alternative is to have two observations over a period of time. With that, the dyad is not so important because you're measuring change, then you only have to figure out what that change is due to. Is it due to neurodegenerative mental illness or is it due to progressive systemic illness that's been poorly treated – congestive heart failure, not taking medications, etc, that have impaired cognition.

Fifth point of departure: reliable information about the presence or absence of neurodegenerative dementia. In my view, it cannot currently be obtained by any biological sample that you can collect, as far as I know. However, that may change. I think we sort of did this last year, where I did present some data, and presented some actual methods of collecting cognitive information, which I'm not going to do because I want to stick more to concepts here. But I think you've told me that you're not doing spinal taps on these home visits, and therefore spinal fluid, [audio unclear] the brain that contains the most reliable biomarkers that we have, are not in the mix. There are biomarkers that we think are going to be relatively informative, and what I presented last year was isoprostanes in both urine and blood, so you can potentially collect those samples.

The only point I wanted to make was that it may be possible, in the not-to-distant future, to obtain biologically-based biomarker data, from samples that can be easily obtained, other than blood and urine, which include hair and fingernails. And I just want to go back to the isoprostane issue there, because this, as I talked about it last year and don't want to go into the details, and certainly not into the biochemistry of it, but it is a very stable marker of oxidative damage to the brain, which may have some specific relationships to certain of the neuro-degenerative dementing illnesses. Because it's stable, we have been able to demonstrate that it can be detected in urine and blood relatively easily. However, practically, it can and has been measured in breath, which is not what I would suggest, but I was sort of surprised to find that out. My biochemist has told that there's no reason in the world, assuming you can [audio unclear] it, that it can't be detected in hair and fingernails. So the only point I'm making here is while we don't have a reliable biomarker, certainly if you're going to collect urine, that would be great, love to have about 500 microliters of it, but if you're going to collect hair and/or fingernails, I'm not sure which is more invasive, some people have more of one than of the other. But in any event, to store that, with the idea of that if the techniques of solublizing it, and therefore have isoprostane measurements made, is not unreasonable. It would be a good thing to do proactively, where you might then come back a few years from now and say 'gee, weren't we smart? We did this, and sort of knew it all along'. And if you're not smart, or if it doesn't work out, you just don't mention it and nobody knows, so no big deal one way or another.

The sixth point is the cognitive information in the absence of either functional information, this is how well they're performing, or change information is still helpful, it's just not as helpful. And there are methods of trying to adjust it, there is adjusted sensitivity and specificity of the cut points, when that's the only information you have. The exact same thing can be said of functional information, that having the functional information, from my standpoint, is actually a little bit better than having the cognitive information. If you told me I could only have one, I would take the functional, and then impute the cognitive from that, rather than the other way around, because it's not confounded so much by education performance and motivation. But having only functional information is not as good as having both.

Waite: What do you mean by functional?

Clark: Functional information is simply a description, and this can be standardized fairly easily and obtained with almost no time required by the interviewer of what the level of performance there are in some very standard activities. So it's basic activity [audio unclear] living, but you don't use, it's not like can they go to the toilet by themselves, can they dress, etc., it's how is their memory, and then there are six descriptors of varying severity: no memory problems, has a little bit of memory problems remembering lists but doesn't interfere with daily function, has memory problems holding on to new information, and goes all the way down to can't even remember the basic, what's going on. But you have to get it from somebody else. I'll get down to strategies as soon as I just finish the last couple of items on the points of departure. But it's very doable, it's very efficient, and if done in a standardized manner, quite reliable, and certainly repeatable. Seventh, functional information alone is useful, but not as useful as the both of them, and cognitive information alone, the same. I guess the last thing is just a reiteration that change is THE most sensitive, and of course, if you can get all of that information, then you really have a lot. Now, what are some strategies for doing that? Well, the standard strategy for collecting cognitive information is to ask some simple questions that key on the major changes in cognition in the dementing illnesses. And since the major dementing illness is Alzheimer's disease, accounting for about 70% of diagnosable dementia in the community – you know it's age-related of course, but it's 70% certainly around age 75 – the standard methodology is to take a look at the domains that are most affected by Alzheimer's disease. That, of course, first and foremost, is memory. So first, a brief memory test, second is language, and third is the ability to – the neurological term is praxis – where the ability to do motor tasks that involve some sort of mental manipulation. So you copy a design, you draw a clock, those sort of things. You want to touch on all three of those domains, and what the field has been trying to do, since Fulstein invented this component of it, is to determine how few questions can you ask and get reliable information. So in 1975 Fulstein did the mini-mental stat exam, which is the first shrinking of the two-hour neuropsychological battery. A couple of years ago, Fulstein published the micro mini-mental status exam. But the issue has always just been getting down to the fewest questions that you possibly can do. It is possible, on current technology, to essentially get it down to one domain: this is visual memory, which can be computerized on a notebook computer. The subject themselves, once instructed, can actually do the test in about 8 minutes with a high degree of reliability, and it correlates very well with most of the other cognitive domains. When you actually measure all four, or all three of those major domains, most of it loads on

memory. So that is doable. What you really want to do is maximize the efficiency of your data-gathering, right? You could easily envision a strategy where, since you have to interview two people, for one of whom you're mostly interested in the cognitive performance, and the other in functional observations, that those interviews could occur at the same time. Also, because you don't necessarily want to have somebody, you absolutely don't want someone talking about the functional abilities – let's use the model of their spouse – while their spouse is there. You can set that up, you don't have to interview them in separate rooms, as long as there's no verbal reporting, you can have it done, you can have both of them operating on notebook computers, one doing a simple visual cognitive memory test, and the other simply choosing answers on a five-question multiple choice questionnaire with instructions. They're both working silently at the same time, collecting that data, that then you can download into a database without any transcription of data error problems, because you're not going from paper to database. The third thing to think about on the functional assessments, at least what my urging would be, is that you be flexible, consider being as flexible about it as you can be, in other words you don't have to gather the data the same way each time. The important thing is gathering the same questions in the same order, with the same instructions. But for people who don't want to answer the questions at home, there's certainly no reason why you couldn't leave them with a piece of paper, have them circle the answers and mail it back, have them call up, and have your interviewer receive that phone call and get that information over the phone, have them call up, and have an automated phone system, so you can even take out the receiver human element of it if you want it, and be as flexible about this as your creativity will allow. And that's probably it. Those are thoughts on strategy, and some thoughts on things that you need to focus on, or least that you need to be aware of, in collecting markers of neuro-degenerative demented related cognitive impairment.

Wolfson: I'm just curious to understand, what is the purpose of measuring the cognitive function? Is it to assess capacity to think, is it to make a diagnosis of dementia, or is it to just identify people who have cognitive impairment along with other things like that? Because I think that the strategies are a little bit different depending.…

Clark: Absolutely.

Wolfson: You've talked primary about diagnosis of dementia, not so much capacity, the [audio unclear] category that was picked up in the NHANES study on aging –

Clark: What I was actually focusing on was that the use of a cognitive measure, or use of any community-based information to identify the presence of a dementing illness or a person who has a high probability of progressing to a dementing illness, not diagnosing the dementing illness itself.

Wolfson: But in order to identify the [audio unclear]

Clark: If you only want to identify who has cognitive impairment, I don't think you need to do anything more than administer a brief standardized cognitive test for which you have age and education and ethnic and racial – all those things are important -- norms that'll give you the prevalence of cognitive impairment, at least based on norms, but I'm not sure how useful that is. It isn't going to tell you about disease.

Unidentified Speaker: My question is what are you going to be measuring cognitive function for?

Waite: Well, in the NSHAP study, if you think about health, it has multiple dimensions: physical health, emotional well-being, cognitive function. It's a measure of health in one dimension. So we are interested in the development of dementing diseases and cognitive impairment, but it's a measure of health. How healthy are you in your cognitive function?

Unidentified Speaker: The other possibility that I would hypothesize and be interested to see is that I would imagine that some people could perform better if, say, a widow acquired a partner in the interval during the time that we interview them. So an alone person may be vulnerable to some cognitive decline, but then when partnered with somebody who is equal or better in their cognitive function, they actually improved in an interval of time. And I think that that's a hypothesis that we're interested at looking at in the longitudinal data. I'm just not sure that we have an instrument sensitive enough.

Clark: Well, that's the problem. I think that what you're saying might be true, but there's no instrument that I know of that you could use in the field that would detect that level of difference, because I don't think that those are going to be robust levels of difference. One other point that I forgot to make is that as we are reminded by the lunchtime talk by the FBI, there's a lot of very useful information and observation that if you just build in a couple of questions that your interviewer answered by their observation, it would provide good supporting evidence, or at least evidence that could be built into the models of the data analysis. For example, if the interviewer walks into a house and there are two people that they're interviewing, it's a typical spouse model -- sit down, turn to one of the spouses and starts collecting information, and says to them, okay, what's your date of birth? The next ten seconds can give you an awful lot of information.

There are a couple of responses. One is the date of birth trips off the tongue with no problems, that's a very useful observation. The second is there's a pause that exceeds 5 or 6 or 7 seconds, that's a useful observation. From itself, it doesn't tell you a lot. The third possibility is that the spouse turns to their spouse when you ask the question. That's a very powerful observation. And the fourth is, that the spouse says "a long time ago." That's a very powerful observation. None of that will tell you whether they're demented or not, but used in conjunction with other information that you may pick up on in cognitive testing, or on functional observations that the spouse provides for each other, that may help you determine how reliable those observations are.

Studying Cognitive Function in the Population Setting: The Challenges

Speaker: Brenda Plassman

Studying Cognitive Function in the Population Setting:

The Challenges

Chicago Workshop on Biomarker Collection in Population-Based Health Research

June 10, 2004
University of Chicago

Brenda L. Plassman, PhD
Duke University Medical Center

Before selecting measures to study cognition in the population setting, one must answer the question that Christina Wolfson asked: 'What do you want to measure?' The answer to this question will guide the choice of cognitive instruments. Chris Clark mentioned a number of instruments that could be used and there are numerous other candidate instruments also. I think it is important to note that individual biases, based on familiarity with a measure, often drive the decision regarding the choice of a cognitive measure. However, I think it is important to base such decisions on information derived from an empirical assessment of a given test.

As a starting point, I think it is important to operationalize the criteria for the different cognitive states of interest. Once the criteria are clearly spelled out, it is easier to identify the cognitive domains that need to be assessed and the necessary psychometric properties for the measures. If you just want to delineate between the two broad categories of 'dementia' and 'not demented,' the choice of cognitive instruments would likely be different than if you wanted to also diagnose mildly impaired cognition.

Operationalize Definitions for Cognitive States of Interest

- Dementia and/or subtypes of dementia
- Mildly impaired cognition
- Normal baseline cognition

Challenges

- Selecting appropriate psychometric instruments:
  - For range of cognitive domains necessary to meet diagnostic criteria (e.g. verbal and non-verbal memory, orientation, language, mental processing)
  - That have norms for age, gender, education, race/ethnicity – that were developed in comparable sample
  - That equal minimum test battery with sufficient specificity and sensitivity for targeted diagnostic groups in sample of interest

Selecting the actual cognitive instruments will depend on the cognitive states of interest (i.e. dementia, mild impairment and normal cognition), the range of cognitive domains to be assessed, the demographic characteristics of the sample, and the availability of appropriate test norms for that sample. Unfortunately, at this point in time the available normative data for most cognitive measures is rather limited for minority populations and very old individuals.

The importance of appropriate norms is exemplified in our work. I work on three large epidemiological studies of dementia. In all three studies, we administer the Mini-Mental Status Examination (MMSE). This measure has a maximum of 30 points on it and is routinely used as a general screen for dementia. The published standard cut point for dementia is a score of 23. However, in one of the studies I work on, it is not uncommon for us to diagnose dementia in individuals with an MMSE score of 27. This happens to be a fairly homogenous group of high-functioning individuals who are residence of a single county. In another of our studies, we have diagnosed individuals with an MMSE score as low as 15 as 'not demented.' This latter group is a cohort that is representative of the national U.S. elderly. Comparison of performance in these two samples provides a basic example of how standard test norms may not be applicable to a non-clinical or epidemiological sample.

Challenges (cont.)

- Follow subjects longitudinally to clarify cognitive status and etiology of any pathological conditions

The last point on this slide emphasizes the need to develop the minimum test battery. Recently, the Alzheimer Disease Centers (ADC's) have convened a panel of individuals to compile a minimum test battery for the diagnosis of dementia. I believe the battery, as proposed, is about an hour long.

As Chris Clark mentioned, it is important to follow individuals longitudinally. If that is not possible, it is important to be able to determine whether performance on a specific cognitive measure represents a change from the individual's baseline. In some situations, this is fairly easy to determine. For example, if you're assessing a retired physician and he or she scores poorly on a test, it is likely a change from their baseline. However, in epidemiological studies, participants come from a range of educational backgrounds and thus it is often more difficult to determine whether current performance represents a change from baseline.

Challenges (cont.)

- Measuring other factors that may affect cognitive and functional performance
  - Sensory impairment - measure of vision and hearing and a subjective rating by examiner to estimate extent the impairment affected performance on psychometric tests
  - Presence and severity of medical conditions (self/proxy report vs diagnostic measures)
  - Medications

Many factors can affect cognition. In population studies, it is important to assess these other factors. For example, sensory impairment – hearing and vision – is a big issue in the very elderly. Other issues are general frailty, medical illness, and medications.

Challenges (cont.)

- Collecting subjective information from knowledgeable proxy about subject's level of functioning in daily activities, including an estimation of the extent to which any decline in function is due to cognitive or physical impairment

In population studies, it is particularly important to interview a knowledgeable informant for the participant to collect information on the participant's level of functioning in daily activities.

How often are medical illness and sensory impairment the main cause of mildly impaired cognition ?

- Among 242 subjects diagnosed with mildly impaired cognition in an epidemiological elderly sample:
  - 22% - impairment attributed to medical illness and sensory impairment
  - 19% of the medical illness and sensory impairment group have sensory impairment severe enough to be considered one of the major causes of cognitive impairment

Cognitive impairment that does not meet criteria for dementia likely has many etiologies - not all mild cognitive impairment is prodromal AD. This slide shows an example from some of our work. Among a group of a couple hundred individuals who were part of a national sample, for about one-fifth of those diagnosed as 'cognitively impaired but not demented,' the only cause for impairment appeared to be a medical illness or sensory impairment. The medical illnesses included such conditions as COPD, diabetes, congestive heart failure, and other cardiovascular risk factors. For one-fifth of this group, the cognitive impairment appeared to be due solely to sensory impairment. That is just a brief overview of some of the issues in assessing cognition in population studies.

Unidentified Speaker: I attribute all my success to my psychology colleague, so that's not a criticism. 19% or 22%, cognitive impairment is due to something else –

Plassman: In about 20% of the 'cognitively impaired not demented,' the impairment appears to be due to medical illness or sensory impairment.

Unidentified Speaker: The problem, however, that I just want to raise when attributing cognitive impairment to medical problems is in some cases in some ways it is almost a Catch-22 situation, because you can perform poorly on a cognitive test because your medical problem is poorly treated – for example, you haven't taken your blood pressure medicine, you haven't taken your asthma medicine, you haven't taken your insulin – or you can not take your asthma medicine, blood pressure medicine, and insulin because you're becoming cognitively impaired. And it can be very difficult to sort out which is the driving problem. So I'm always a little suspicious whether that 22% would no longer be cognitively impaired if you took away those medical problems.

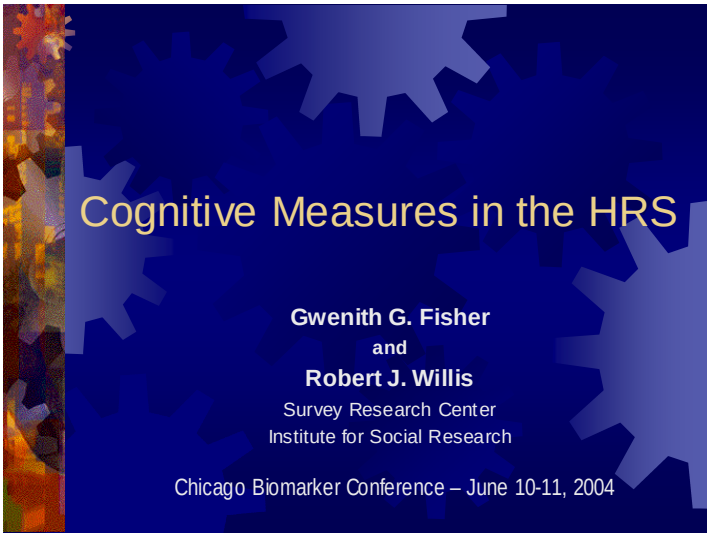
Plassman: I agree that it is very difficult to tell which is the real problem. In population studies, it is even more difficult to sort it out because the study participants are not presenting at the doctor's office with a complaint of cognitive impairment. In the proposed study, individuals are going to be assessed in their homes. From my perspective, this just reinforces the need to collect as much information from as many sources as possible to try to determine the individual's real level of cognitive and functional performance.

Unidentified Speaker: Let me just make one more point, then we'll shut up about this. If you said medical problems can reduce cognitive performance, even below norms, I would say fine, live with that. But I really have to say that in a couple of thousand patients with dementia that I cannot recall a patient who was flat-out clinically demented where the problem was due to a reversible medical problem, not talking about AIDS and [audio unclear], but had a progressive dementing picture where it was due to diabetes, asthma, congestive heart failure, circulo-pulmonary disease. I think you can affect cognitive performance, but I'm not sure you can drive somebody to become demented just through lack of medical care.

Plassman: We would generally not call an individual "demented" in situations where the sole cause of cognitive and functional impairment was thought to be due to medical illness. However, other groups may disagree with us on this point. But these individuals are functionally and cognitively impaired and the impairment does not appear to be solely due to physical impairment.

Cognitive Measures in the HRS

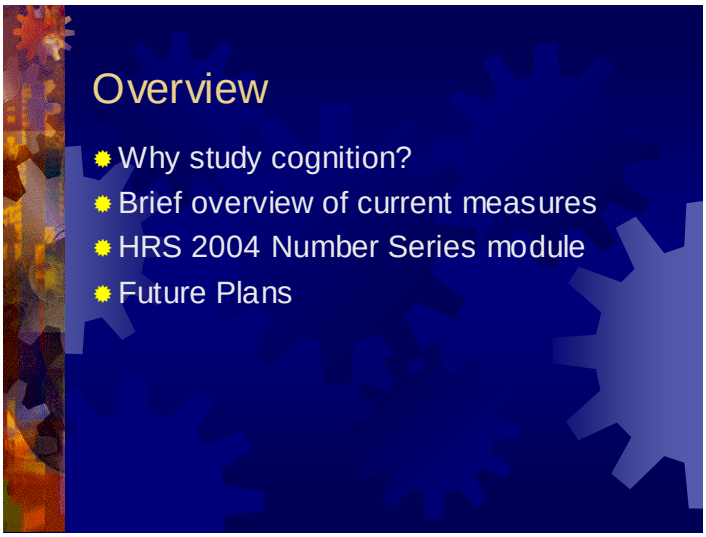
Speaker: Robert Willis



I'm actually stealing these slides from one of the people on the National Health and Retirement Study staff, Gwen Fisher, who put a very nice presentation together for our staff about cognitive measures in the HRS. Let me just give a two-minute description of the HRS. It is a nationally representative longitudinal panel where people are enrolled in the study when they enter their early fifties, and they're followed longitudinally, every other year, from the time they come into the study until they die. That is the overall design. The study's been going on since 1992. It began with a study of the head sub-sample of the HRS, that began with people who were seventy years of age and up in 1993, and another sample of people who were 51 to 61 in 1992. So a lot of the longitudinal information comes from those two sub-samples. But in the long run, it has this population's representative character.

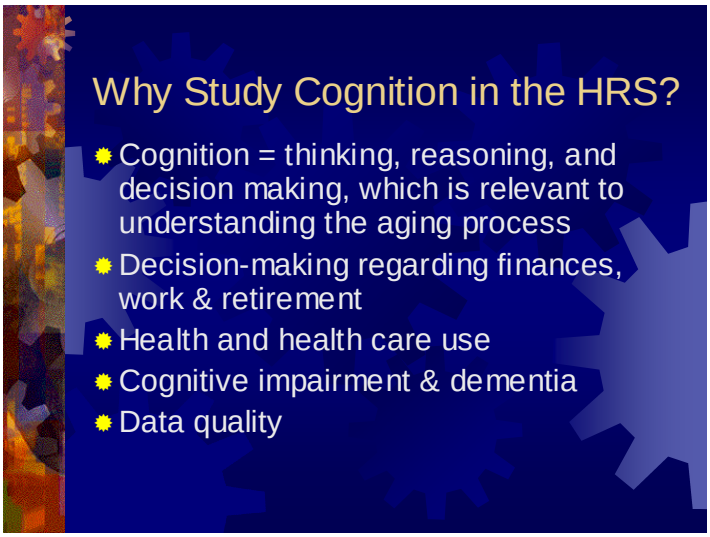
The other thing I'll say about it is that the goals of the HRS are really quite wide. It was created as a public-use dataset for the research community that deals with the wide spectrum of issues facing not only individuals as they age, but also society as it gets older. So we know, for instance, that the new cohort, which we're bringing in 2004 to keep us representative of people over age 50, are people who were born in 1948 to 1953. Those are people who are at the leading edge of the baby boom, and they are the people who, if they decided to retire early and had a lot of health problems, are going to quickly bankrupt the Medicare system and the social security system.

On the other hand, if they decide to work a long time and stay healthy, they could at least postpone, if not avoid, some of these issues. So they're really a critical group and it is important to understand their behavior from a wide variety of points of view. Are they well-prepared for retirement? What kinds of assets do they have? What kind of savings have they accumulated? What's the influence of public and private pension programs on their retirement behavior? And so on. Then later, we want to understand how they cope with health problems, how they interact with their family in a wide variety of things.

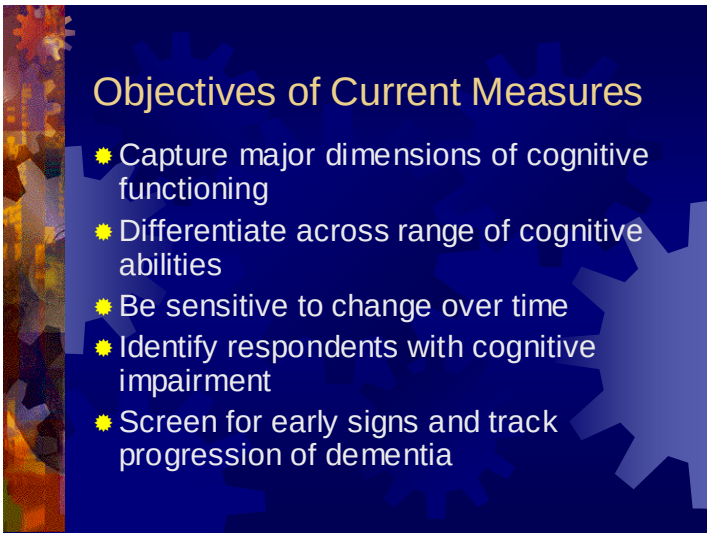


Why study cognition? In the context of the goals that I talked about, cognition is really quite critical. The world is becoming ever more complex for older people, particularly for older people. They're facing financial decisions, dealing with portfolio issues, the study of which has won Nobel Prizes, including ones here at the University of Chicago. They are studying medical care problems, they're trying to choose drugs and so on. These are very complicated decisions. These decisions may be individual decisions. However, they are involved very often with families. So one needs to think about not just individuals, but also family interactions.

I want to give you a brief overview of the HRS measures that we have, and then suggest to you that we actually are reengineering those measures, with a major study that we're hoping to have in place on the HRS in 2008, and I'll describe a little bit of what's going on there. That's what I'm mainly going to do. We have a number-series module that I'll describe, which is kind of a precursor of that reengineering, and, then I'll talk about the future plans.

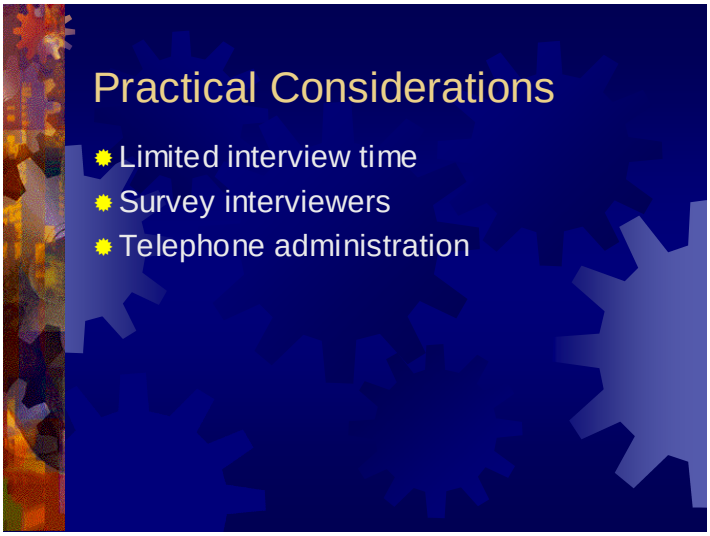


So why study cognition in the HRS? I think for this group it's not too surprising. We think that cognition involves thinking, reasoning, and decision-making, which is relevant to understanding the aging process. Economists are coming to the realization that we have traditionally operated on models that assume unbounded rationality, where people can deal with a problem, no matter how hard, and make optimal decisions. Increasingly, we think that that's not really the right model, and we need to understand limited abilities to do these things. Decision-making regarding finances, work, and retirement I have already talked about. Health and health-care use are also affected by cognitive impairment and dementia. As I said, all of these decisions involve complexity, including the last part. A person who's becoming impaired and demented, for one thing, has to look toward the future and think about, well, what would I myself do were that to happen to me? And secondly, the person involved is embedded in a family and in a community and so forth, and there are a lot of very complicated things that those actors have to do when this takes place. All of these areas involve complex decision-making, and that's actually, in the future of the Health and Retirement study, going to be an increasingly important theme on decision-making. Data quality: how well we can understand these things does depend a lot on what kind of quality of data that we can bring to bear on this.



I want to talk about the current measures and really an assessment of the strengths and also the quite significant weaknesses of these measures that we want to try and repair. We'd like to have measures that capture the major dimensions of cognitive functioning. We'd like to differentiate across a range of cognitive abilities. We'd like to be sensitive to change over time, not only for the purposes of diagnosing dementia, but also for other purposes.

Psychologists who have been researching cognitive aging have recognized that there are a number of trajectories to various components of ability which are not related to disease that may or may not -- we really don't know -- impact a number of these kinds of decisions we're talking about. We want to identify respondents with cognitive impairment. For that purpose, we've mounted a major study that Brenda and Bob Wallace and Ken Langa are all involved with, called the Aging and Demographic Memory Study, or the ADMS Study, which is attempting to get a clinical measure of dementia, and the sampling frame is imbedded in the HRS, so one of our purposes is to try to understand the connection between short measures that could be conducted in a survey and clinical data that's as good as we can get. And if we can discover good measures of that sort, we can, I think, radically extend the capacity of major social surveys, rather than special surveys, to take into account dementia. So that's a project that you'll hear more about in a bit. We also want to screen for early signs and track progression of dementia.



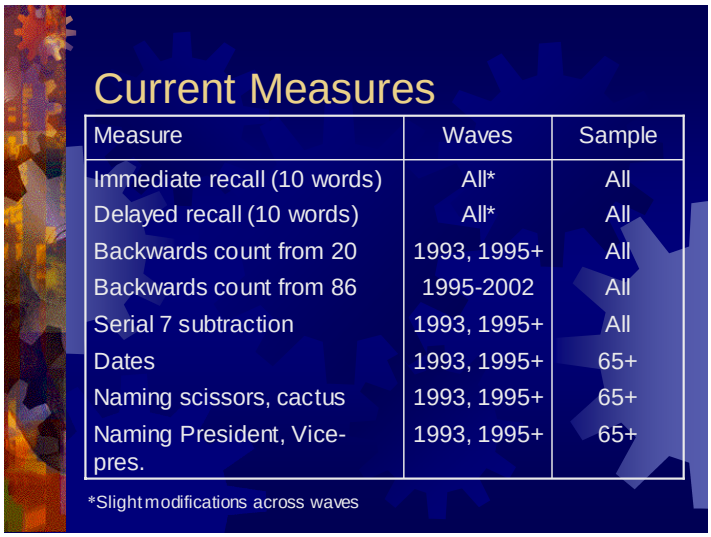
Practical considerations. Limited interview time: The HRS tries to cover a great deal of territory, and in fact it's major strength is the fact that we cover major interacting variables that are normally in the domains of economics, of social demographers, of psychologists, of medical people, of health services researchers, and so on. We really do need to optimize, and find optimum ways to measure what's most important in each of these domains and allow them to interact amongst one another.

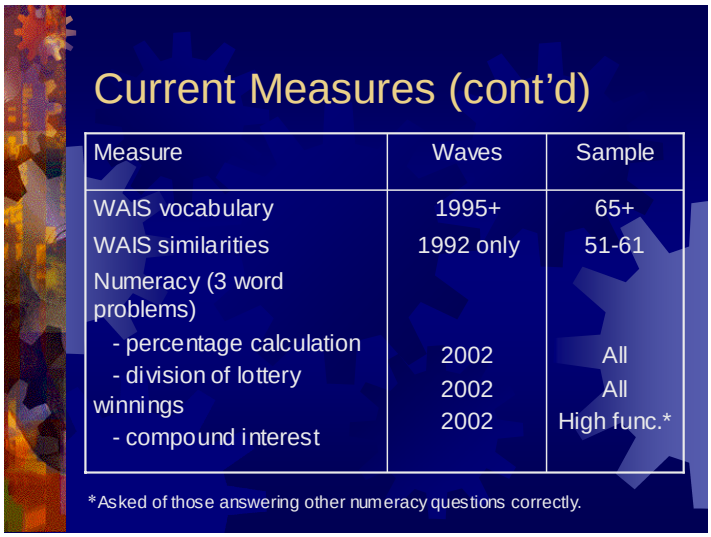
Survey interviewers: We have survey interviewers who are relatively generally trained. In the ADMS Study, we have a nurse and a psychometric technician who visit a home. That's a very expensive proposition. We can't in general do that kind of thing, so we want methods that can be utilized by more general kinds of survey interviewers.

Telephone administration: HRS is a survey that uses mixed modes. We use a personal interview at base line. We then normally switch to telephone administration in every subsequent even-numbered year, until a person gets to 80, and at that point, people are normally switched back to personal. We also have discovered that we can fruitfully use, in odd-numbered years, mail surveys to collect a variety of kinds of measures. Those are self-administered surveys that are obtained by mail.

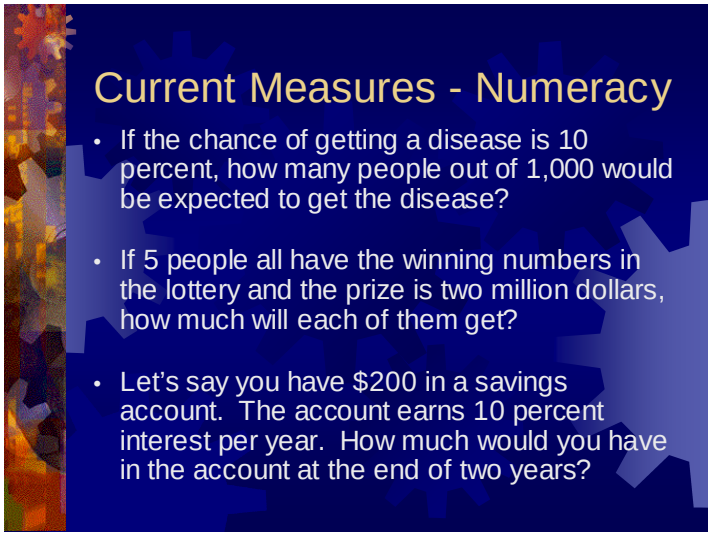
We also have a project, it's a collaborative project between Michigan and RAND, that's dealing with Internet interviewing in this older population, and looking for mode effects and selection effects in that context. In that context, a lot of the cognitive measures have been developed in a personal interviewing mode, and the properties of these measures are not well understood in a mixed-mode setting. That's one of the open questions.

Here are some of the indicators. We have immediate recall and delayed recall, we have a backwards count from 20, a backwards count from 86, serial 7 subtraction. Incidentally, it turns out that the second of those subtractions has quite a bit of power. A very large number of people can subtract 7 once, but it starts to discriminate at the second one. And then the screening items for dementia, involving naming the vice-president and scissors and cactus and so forth.

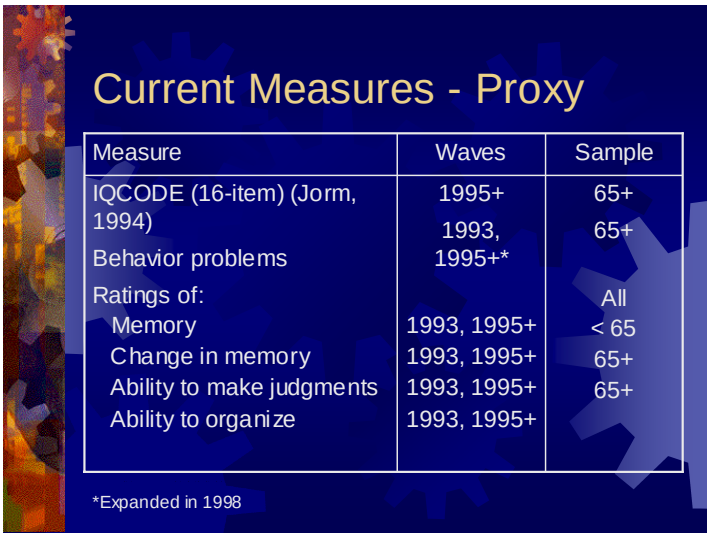




We also have some baseline measures of WAIS vocabulary, WAIS similarities, we have some numeracy (three word problems). We've done some experimental modules, which is actually another method that we use to collect data, which is to try out new items. And we've had some percentage calculations, divisional lottery winnings, and actually calculation of compound interest, which is an issue that's key for economic reasoning and is actually a task that's pretty difficult for a lot of people.



We have some current numeracy measures. Peter Ubel, a doctor who deals with medical decision-making at the University of Michigan, has put some of these on, and we've been trying to get at these. You can see some very simple questions of this sort. These are some examples of those questions, this last one being one that's not asked to people who flunk the first couple.

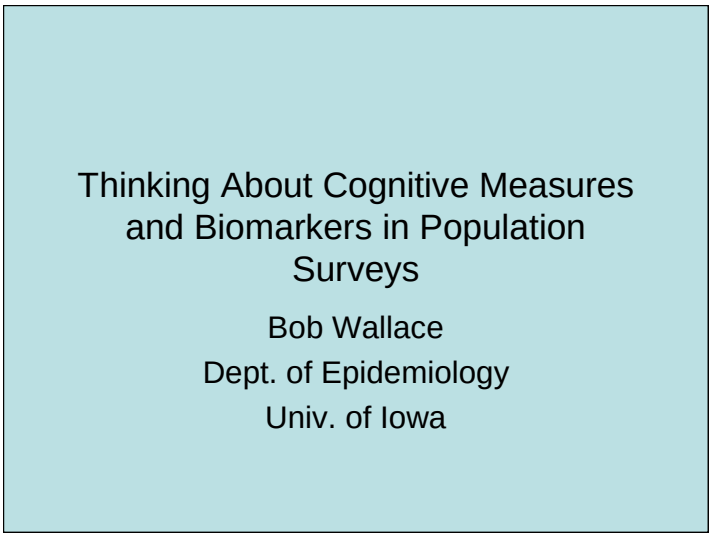


We have a proxy design for the HRS. At a point at which people cannot respond for themselves, we use proxy measures. The original design for the head study was proposed as having proxies for everyone, but the reviewers decided that that was too expensive. So what we have is proxies for people who are unable to do the interview themselves. We use the proxy to get some evidence on dementia using the items listed in the slide. Rethinking the cognitive measurement in the HRS, we have an oversight group that advises NIA and works with the project that decided to commission a content review of the HRS midway through our funding cycle. So the content review is not connected to a funding review, and I think it was a really successful enterprise in which we got experts in a variety of ways -- some of whom have had a connection as users of the data, and others of whom had not -- to do a review. One of the areas that was reviewed was by Margie Lachman and Avon Spiro, on the cognitive measures, and they applied in many strengths of the current cognitive measures, but they reviewed and highlighted a desire to assess a broader range of abilities to differentiate more effectively among higher levels of cognition, which I think which are the ones that we economists surmise are involved with a number of these financial planning tasks and so forth. We decided to take some directions.

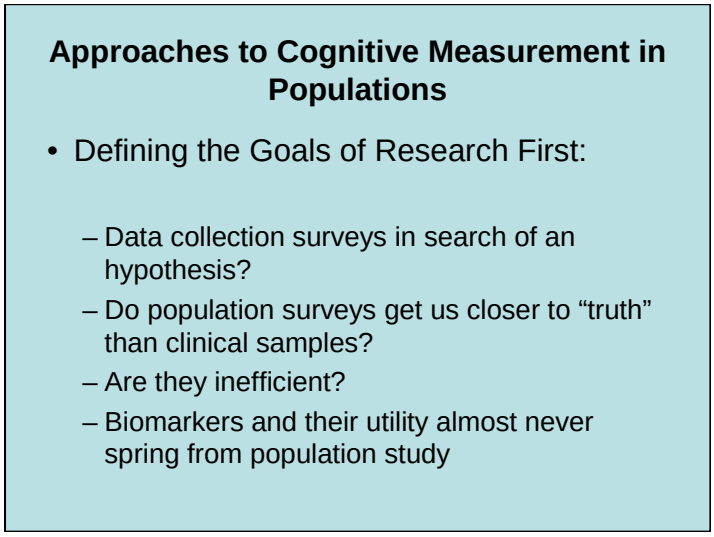
Thinking About Cognitive Measures and Biomarkers in Population Surveys:

Speaker: Bob Wallace

In this brief presentation I will offer some personal thoughts on biomarkers in population surveys and cognitive tests, and then go on to answer Ken Langa's four questions.



I want to start with a cautionary note about biomarkers, particularly those related to cognitive performance. I'm not opposed to searching in population studies for new biomarkers; I'm basically in favor of that. This notion may seem simplistic to you, but I have seen a lot of scurrying around in search of biomarkers without really considering the stage of research and whether the investigators really have an hypothesis. Some of the issues appear in the following slide:

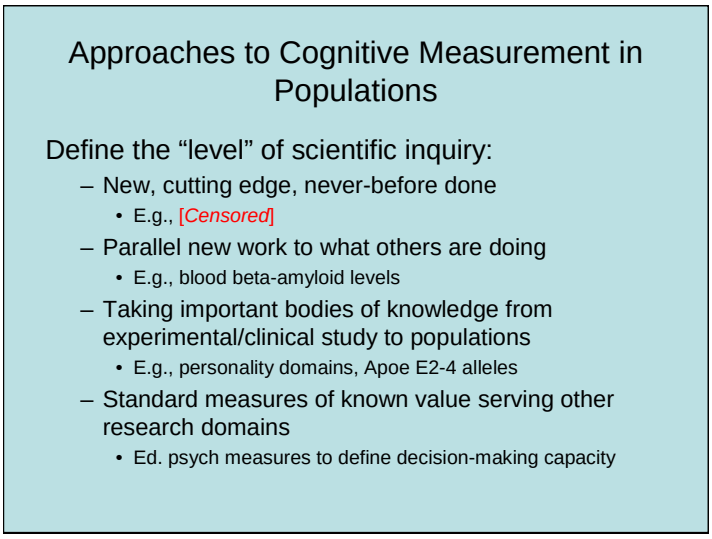


Sometimes this scurrying can be quite problematic, because there have been many data collection efforts, including surveys, in search of an hypothesis. My argument is that populations are not necessarily the best place for conducting a search for clues to hypotheses. Chris Clark said this earlier today in a somewhat different way, and he was right to suggest that there was substantial heterogeneity (variation) in general, representative populations. When one is attempting to prove a principle of an association between a marker and a phenomenon, or a marker and a disease, a population may be the last place I'd want to go. In many instances, clinical or volunteer samples may actually get us closer to the truth in seeking biomarker associations than populations do, because of participant variance, small sample sizes for many diseases, or physiological phenotypes.

There is also an efficiency argument here. Do you know of any biomarker-disease associations that were discovered in a population first? One only has to be reminded of cholesterol, hypertension, blood sugar, ponderosity, and so on, all with a history of discovery through clinical or pathology research.

The remainder of this cautionary note is that you should know where you are in terms of the how advanced the level of scientific inquiry is in general when you are exploring biomarkers in your own

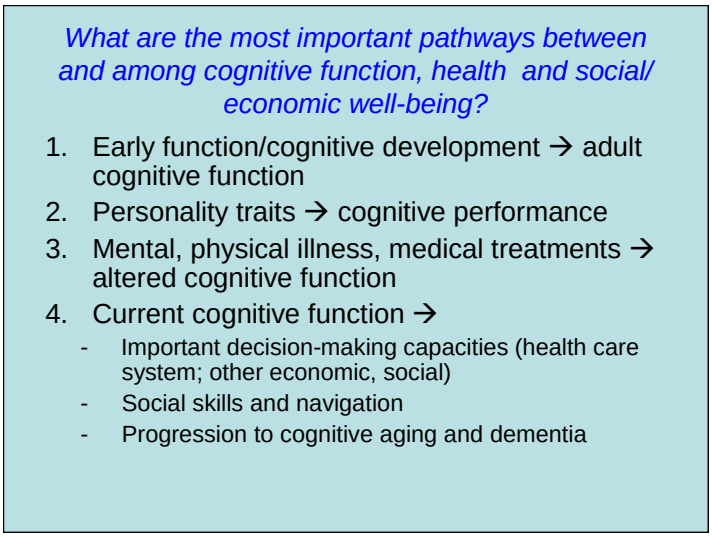
study. The next slide shows some of these levels just for exposition; these levels are not necessarily ordinal, but nonetheless represent different ways to consider which biomarkers should be selected for measurement and discovery.



The first would be new, cutting-edge, previous unexplored markers, based on your own ideas. I was going to give you an example, but I don't want to give away my secret. In this high risk situation, you need resources, and these new ideas are often difficult to get past peer reviewers. The next level in the slide would be to do newer work of an already hypothesized but not proven marker, in parallel with what other investigators are doing. This is important to do and likely to have better outcomes. For example, I had the idea but not the means, along with several other investigators, that blood beta-amyloid levels in populations might predict the onset of Alzheimer's Disease. This is a protein deposited in the brain that is likely to have a role in the pathogenesis of that disease. Because I didn't have the time or funds, it was left to others to determine the utility of this measure. The work on this biomarker is yet incomplete, but several groups are working on it.

Another level of inquiry is to take things that are known from clinical research and bring them into the population study. This is probably the best, or at least the most efficient, use of population studies. Then, finally, as the slide suggests, one can take biomarkers that are proven in one arena and explore other ones. As Bob Willis said earlier, one of the things we're doing with the Woodcock-Johnson educational psychology battery, a three hour process when administered in all its glory, is to explore the role of these measures in personal decision-making.

Here are the answers to all of the four questions – in any case, my answers – and some directions in which I'd like to see this field go. I'm very much an adherent of the life cycle approach to the study of diseases in older people. So to the question of the important pathways between cognitive function and health, the most important to me is to follow functional and cognitive development through the lifespan to determine how the forces of cognitive development and intelligence play out in terms of cognitive impairment late in life. Some of you know of the Nuns' Study, where retrospective cognitive evaluation of teens was related to the cognitive status of older nuns 40-60 years later. Not very many other such opportunities exist, but they are extremely important.



I'm also very much interested in how personality traits lead to late-life cognitive function, because these traits lead to behaviors, especially those that are health-related, that partly determine cognitive outcomes. Third, sometimes I think we miss a lot of mental illness and medical treatments and obvious things that likely lead to cognitive change. I'm embarrassed to tell you that when we started HRS, it never occurred to us to ask either the informant or the primary target respondent, 'Did anyone ever tell you that you had Alzheimer's disease?' While doing all of the elegant measures of cognitive status, we forgot to ask initially if a diagnosis had been made. Schizophrenia was already mentioned, and other mental illnesses such as addiction, depression and bi-polar disease. I just want to be certain that, as people get older and we start looking for cognitive decline and impairment and its causes and its trajectory, we will not miss the big things. I'm always afraid that many persons with major mental illness don't get into surveys, and that's a critical survey problem. The fourth pathway, then, would be current cognitive function. Several people, including Bob Willis and Ken Langa, have discussed where we want to go now in HRS in terms of decision-making capacities, social skills, and just navigating modern complex life.

As to what measures of cognitive function are most important, of course it depends on the hypotheses. As I mentioned earlier, one of the most important is records of school performance and abilities. In keeping with my earlier comments, as shown in this slide, any other records that have implications for late-life cognitive performance can be most critical, including clinical psychiatric records—possibly hard to acquire, evidence of sociopathic behavior from school, prison, or other institutional records. Records from the state or federal correctional institutions may be available and if present, can bespeak many exposures and circumstances directly relevant to cognitive function. I'd like to get Department of Justice records and workplace records, and addiction and abusive behavior. If survey time is available, I believe that the boundary of mental illness and cognition is so important that I would use a screening instrument for major mental conditions. There are many such instruments available, such as the "PRIME-MD." Instruments that will diagnose major mental illnesses, such as schizophrenia, bi-polar disease, depression and substance abuse are available, but often time-consuming and limited for that reason. As this slide shows, there are so many important directions for research that are waiting to be done in the population context.

What measures of cognitive function are most important to better study and define these pathways?

1. Early school, social performance, clinical psychologic/ psychiatric records
2. Evidence of sociopathic behavior: school, justice system, workplace, addiction, abuse
3. Screening for important mental illnesses, brain injury, etc.
4. Personality traits, derived from a) scientific consensus, b) twin and other genetic studies
5. Tests of cognitive performance
6. Tests of decision-making and navigation in stereotyped social settings (e.g., health care system)

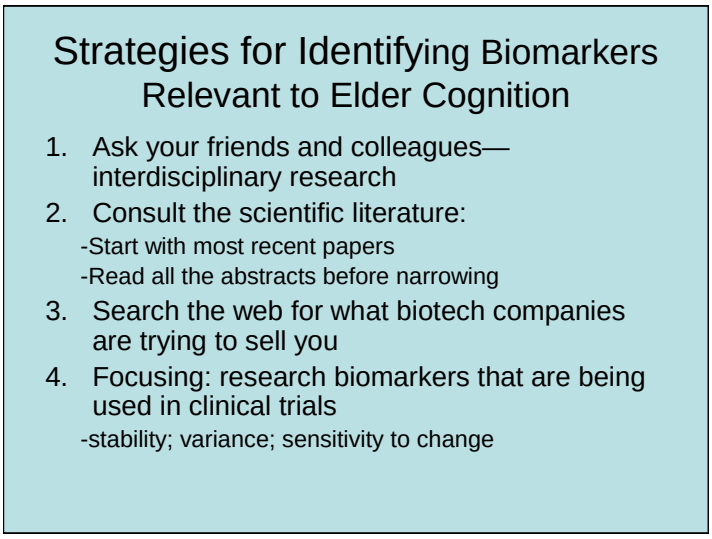
To Dr. Langa's question as to whether there are biomarkers that can predict cognitive performance, the answer is of course, 'yes.' However, to me, the major challenge is to formulate a model that will allow a comprehensive exploration of new markers, and as I mentioned earlier, it may be best to hold population studies until more efficient methods have suggested putative associations.

Is collection of biomarkers for cognitive function feasible and likely to help elucidate…pathways?

- Yes, but short and long-term cohort participation rates unknown
- Possibilities (depending on goals): DNA
  1. Cortisol levels (related to depression, variety of abnormal behaviors)
  2. Sero-markers of possible progression to cognitive impairment: ApoE, inflammation/immune factors, clotting factors, beta-amyloid, alpha-synuclean
  3. Devel.countries: markers of infectious agents; nutrition
  4. Blood pressure recordings

A general problem is that one might expect that many biomarkers that predict general illness and disability will also predict cognitive decrements, which often accompany such illnesses. So the specificity and degree of prediction become very important. Of interest, there is substantial interest in cortisol levels, which have also been related to depression, sociopathies, and substance abuse. So the predictive value of a blood cortisol level for cognitive function must include consideration of these confounders. The same might be said of markers of inflammation and acute phase reactants. There are now so many protein and other chemical moieties that have been related to brain function that there are a large number of exciting candidate markers. Whether they are accessible in the blood and can be accurately measured is another matter. A possible example is alpha-synuclean, which is a protein at the synapse in neural tissue, and it has been measured, and might be useful.

If I were doing developing countries studies, I would add to my search biomarkers of infectious agents, such as HIV, as well as markers of malnutrition. Finally, we should not forget commonly acquired measures such as blood pressure, which has a relation to cognition via several mechanisms.



Ken's third question concerned strategies for identifying biomarkers relevant to cognition, and the slide suggests a few. These are suggestions for the non-biologists: 1) talk to your friends and colleagues—this might include talking to basic and clinical scientists at our local Alzheimer's disease Research Center; search the literature; 2) Go to the edge of the published literature; and read the first hundred abstracts. You'll see that most of the new biomarkers that will appear in population studies next year will be there. They may not be the very newest, but they are the ones that have had some peer review and a higher likelihood of successful application; 3) Search the web for what determinations the research biotech companies are trying to sell you; I get a lot of computer spam in this area, but that's the frontier for reproducible measures if you can afford them; 4) Explore markers of brain function that have been included in clinical trials, particularly of Alzheimer's disease. Often, very innovative measures are included, and usually there is some evidence base for their application. Importantly, brain biomarkers that are altered by the intervention are probably more likely to vary with risk factors and other population variables at hand.

What are the most important challenges to address when studying cognition in population-based studies?

1. Respondent burden/ resistance to cognitive study
2. Acquiring old and possibly sensitive administrative, social, clinical records
3. Validating prior psychiatric, psychological treatments
4. Ethical issues in sensitive measures
5. Selecting best cognitive, personality measures

The last question explores the challenges associated with conducting biomarker studies of cognition. Everyone in this session has of course suggested important challenges, and I have placed my favorites on the last slide. These are all very big problems for us, including the response burden for older participants who may have cognitive impairment, acquiring old clinical records (they are being destroyed at an increasing rate) and validating complex psychiatric illness. The ethical issues, to be discussed next, are paramount. And finally, it is not trivial to select the best cognitive psychological and personality measures, since there seems to be so many versions and brands.

The challenges are great, but the rewards will be as well.

Studying Cognitive Function in the Populations Setting

Speaker: Daniel Brauner

I'm going to talk about something completely different, because I'm not a survey researcher, and I don't know anything about biomarkers. And yet Stacy asked me to come anyway. I'm a clinician who takes care of a lot of people with dementia, and I'm interested in language and dementia, specifically evaluating decision-making capacity. But in reviewing the literature and going to lots of conferences, I think that one thing that becomes really obvious from the studies, especially the longitudinal Scottish study, and many others, is that education and occupation are strongly correlated with the later development of dementia: the more highly-educated you are, the lower the risk of developing dementia. The more skilled occupation you have, the lower the risk of developing dementia. There's always the question of the chicken and the egg here, and we talk a lot about the genetic component, but I think that we've pretty much found that only a small amount of variance is explained by that. Of course other as yet unknown factors will probably come to light in the future and this is where biomarkers may prove useful. However, one thing we know now from the overwhelming evidence from those types of studies is that how you use your brain, from very young ages, affects how your brain ages and the subsequent development of dementia. How it does this, I can't begin to explain. But that it does, I think is something that we know. And I think the important idea from these studies is that education and occupation are really surrogates for the notion of how you use your brain.

People have sort of picked up on this concept a little bit. There's a craze, now, of older people doing crossword puzzles and similar brain exercises in the hopes of staving off dementia I think this craze may be little bit misguided, because the key here is how you use your brain throughout your life. It has something to do with reserve, it has something to do with normal aging, and the line between dementia and normal aging – which, I think it's important remember, is a very conceptual one. These are socially defined diseases to a strong extent. So I'm just brainstorming about looking at other surrogates. These longitudinal surveys are potentially very rich sources to look for other surrogates of how people use their brains during their lives. And trying to figure how to get at that for large numbers of people, so that you can have the power to actually try to understand ways in which, if people could be encouraged or taught to use their brains differently, we may be able to delay the onset of dementia for several years, if not prevent it. This in itself would have tremendous potential for improving the health of large numbers of people.

One of the things that got me interested in this notion was some preliminary data that was presented by Mark Grant at a research-in-progress meeting at the U of C. Looking at retired musicians from the Chicago Symphony Orchestra– a fairly large number of older people among whom you would expect to find a significant number with dementia – he had yet to find a single person with dementia. This was a very preliminary pilot study and the chicken and egg problem persists but it got me excited and thinking about what other kinds of activities or patterns using our brains may have implications for affecting the presence or absence of dementia in later life? What kind of learning or thinking people do that you could measure in these surveys? From some of the things that I've been thinking about, I think it'd be an interesting thing to talk about political affiliation.

Unidentified Speaker: Are you going to disclose your bias here?

Brauner: I have no bias. Other patterns could be how open-minded you are about different aspects of how you engage with the world, with ideas. From religiosity, fundamentalism, going to more liberal views of religion. In terms of your conversations, when you were younger, what kind of things did you talk about, how many turns were in the typical conversations that you would have? How many books have you read? How many books did you read in your twenties or thirties or forties, vs. how much television have you been watching? I think getting at those kinds of notions – and large data sets are available, where you could get some ideas – would be an interesting way of getting at this question. Of course, this doesn't appear to have a lot to do with biomarkers, but perhaps we should start searching for biomarkers that may correlate with these different patterns of using one's brain. I think that this notion has not really been fully explored to the extent that it could be, and I think it could potentially be a worthwhile area. Since there are a bunch of people here doing longitudinal studies, I just wanted to make that pitch. Thank you.

Unidentified Speaker: Dan and I have just started working together on a project, and I've done previous work in literacy and how literacy affects health, but I'm really interested in looking at how literacy is correlated with cognitive function and change over time. We actually had a literacy instrument that was a two to three minute literacy tool in the NSHAP study that got cut in the cutting process. But it seems to me that you're interested in language and its relation to cognition. Literacy measures something different than education status. In fact, they're poorly correlated -- they're not very well correlated at the lower end. So there could be a lot to be learned there. If there are any national longitudinal studies, like HRS, they might be interested in including literacy there.

Unidentified Speaker: I think that Dan pointed out something very important, and that is that early life experiences may be extremely important in these kind of cognition studies. I think there is a literature out there about early musical training being associated with dementia. Certainly when a person learns to read and how many different languages a person learns over their lifespan are all kinds of measures of utilizing brain mannerisms

Unidentified Speaker: I have two comments. One is just a factual one: in the Health and Retirement Study, we have now been doing mail surveys in odd-numbered years that deal with conception, behavior, and time use. And the time component actually gets at a number of the kinds of activities: television watching, and reading. Things that [audio unclear], and those are longitudinal, so that one could see what might happen, say, as a disease condition came along, or somebody went from workplace to retirement, or things of this sort. The second thing I would argue is that there does exist a number of several long-term longitudinal studies: the NLS-79 study that [audio unclear] directs is one that started with people who are 14 to 17 years of age, and follows the children of the female members of that cohort. That cohort is now getting to their early forties, 40 through 48, and as time as goes on, I think, the NLS-79 has gotten

interested in asking 'Do we really want to update the survey to include more health items?' It seems to me that that's a potential resource for people with an interest in health, particularly ones that are interested in long-term development. Another survey is the one that's been going at Michigan since 1968 that follows as kids in the family for example, the children of [audio unclear]. Once again, those are kinds of studies where, by linking health information with a lot of the long-term information [audio unclear] again, the study concentrates on older people, but we have linkages to their social security earnings history, so we can actually know their lifetime incomes, so we ask some things about early childhood retrospectively. So again, there is the potential at least to look at some of these.

Unidentified Speaker: Again, the other part of that someone pointed that in our sample, there are siblings, so there is some control for some of the genetic components that you might see. As well as the PSIP, where you are following families around, see some of that going on. Of course, the component's just really hard to get at.

Unidentified Speaker: It just seems that everyone's just assuming that if you read more, you're going to retain your cognitive health through older age. Is that demonstrated?

Unidentified Speaker: In fact, that's been demonstrated not to work. There was just a very large clinical trial that was published in JAMA, called Activ, that actually did cognitive interventions and had no effect on function down the road.

Unidentified Speaker: That's done much later in life, though.

Unidentified Speaker: I think the point is that these things are happening either in childhood, or in the 20's, 30's, 40's -- the time when you're actively involved in your occupation, when you're being educated.

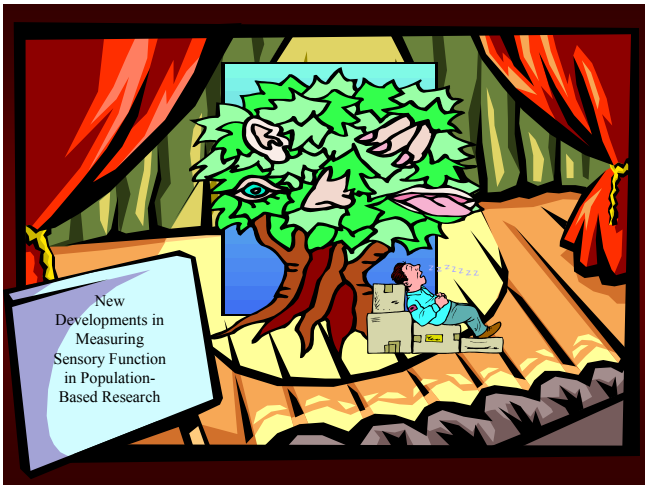
Unidentified Speaker: Do you think that early development affects the expression of pathology or expression of the symptoms of pathology?

Brauner: I would think that it would probably be more the symptoms, but I think that the symptoms basically are the disease, because I think that it has a lot to do with reserve, in that if people develop a larger reserve they could withstand more pathology and still be able to function at a level that's not discernable as disease.

New Developments in Measuring Sensory Function in Population-Based Research (and what can they tell us about health?)

Thinking About Cognitive Measures and Biomarkers in Population Surveys:

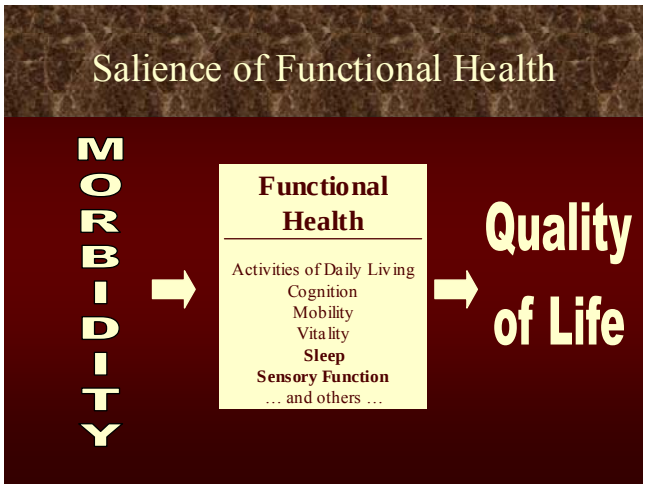
Speaker: Sara Leitsch

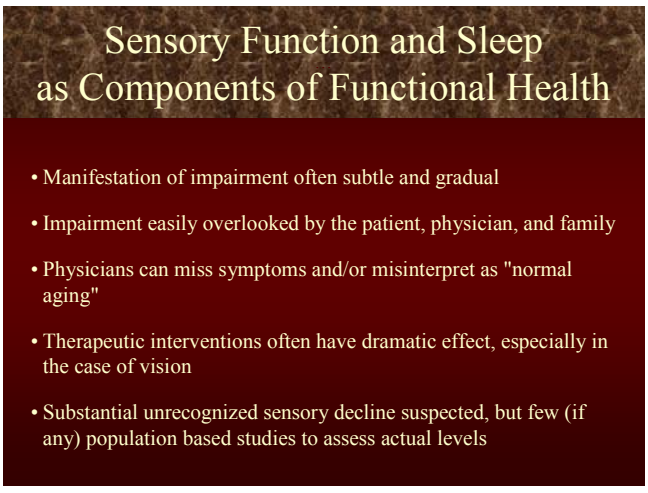


I was asked to moderate this session and to put sensory function and sleep behavior within the context of the health schema about which we've been talking.

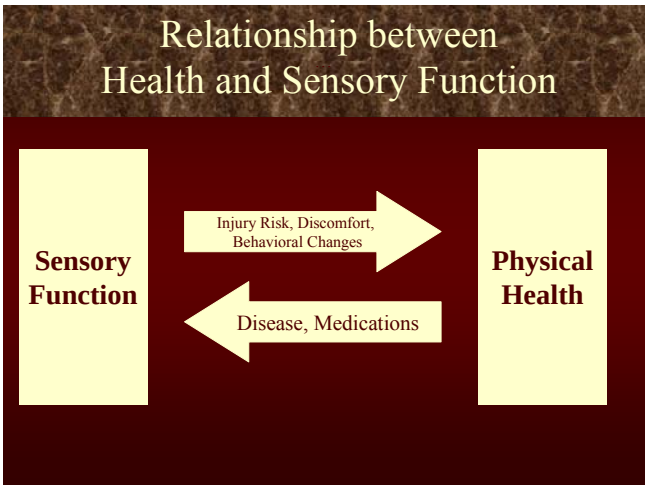
First, I want to emphasize that sensory function and sleep behaviors are part of health. In many cases, they mediate the relationship between morbidity and quality of life.

Functional health is the way in which we experience our world. There are several unique features of sensory function and sleep. The onset can be subtle and gradual, they are often undiagnosed, and they are often unrecognized by the individuals who are experiencing them.



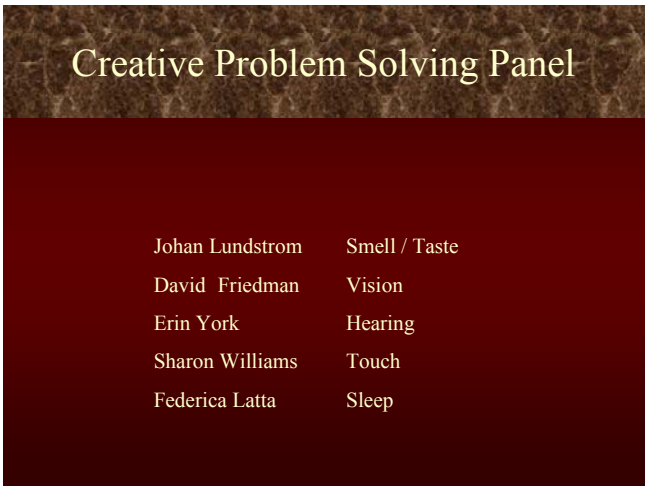


The one thing I do want to emphasize is the reciprocal relationship between sensory function and physical health. Components of physical health, such as morbidity and medications, have detrimental effects on sensory function, but the relationship is reciprocal. Auditory problems can cause headaches and dizziness and difficulty with balance. Sensory dysfunction can increase the risk of injury. If you can't see well, you are much more likely to crash into something when driving. They can also create changes in nutritional behavior, health access, and other health behaviors.



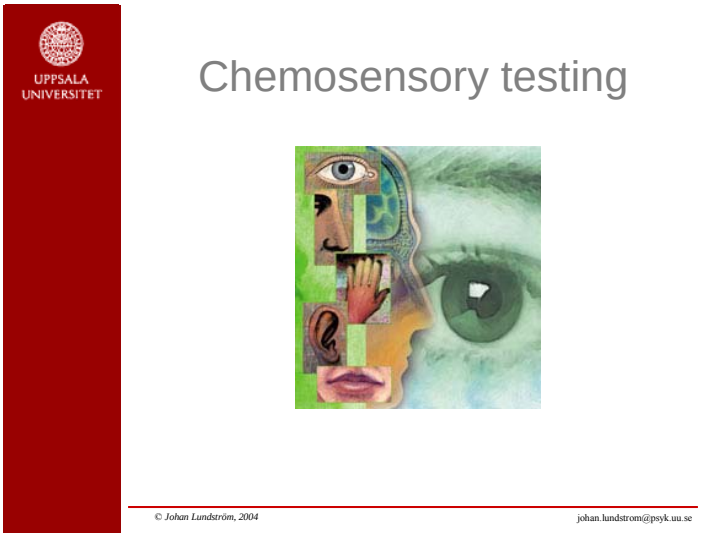
I also wanted to emphasize that sensory and sleep deprivation can result in psychological ramifications. These are some that have been documented in the literature. This is an untapped area. In our pilot study for the NSHAP project, about which you've heard today, we looked at the relationship between perceived olfactory function and satisfaction with sexual life. Women who reported a poor sense of smell were much less satisfied with their sex life.

And now I will turn it over to what I like to refer to as our creative problem solving panel. I've asked all of our panel members to focus on the methodological components of their presentation, because I think that is the most useful for us.



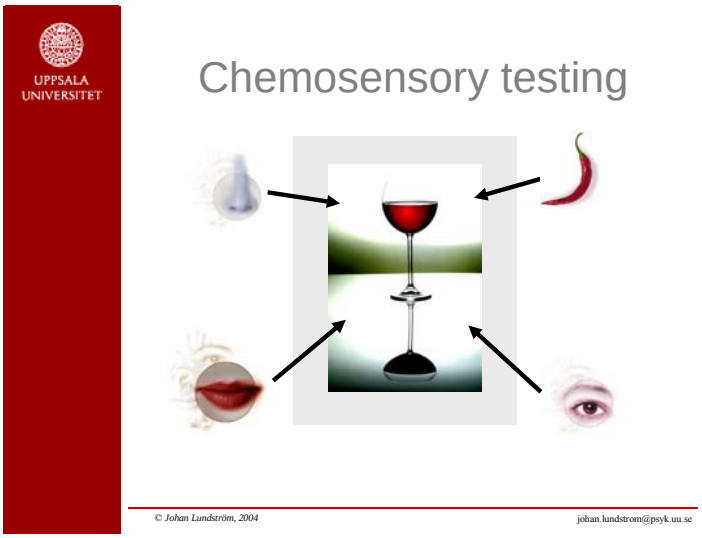
Chemosensory Testing

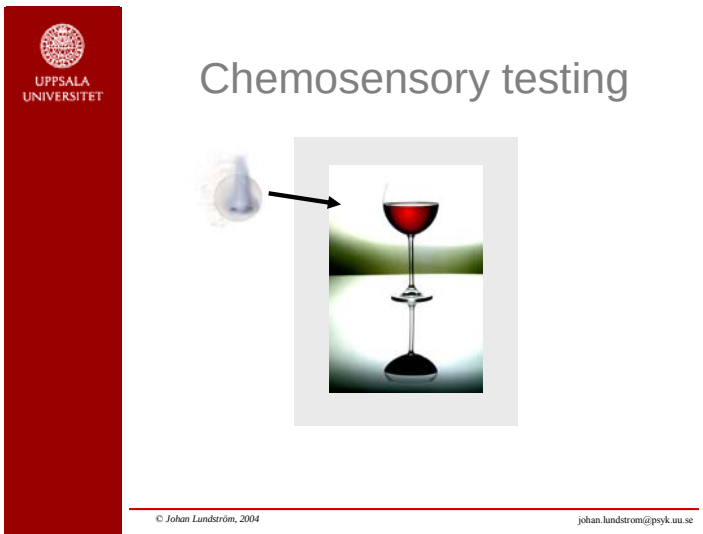
Speaker: Johan Lundstrom



What I'm going to talk about is two of our five senses: olfaction and taste. Further, if I have the time, I will talk about some new ways of measuring sensory functions that are applicable to all the five senses. When those of us in science talk about olfaction and taste we are talking about something completely different than what 99% of the rest of the population talks about when they mention taste and smell.

When you talk about your sense of taste, for example like when you are having a good glass of wine, you are actually talking about a combination of tasting it, smelling it and experiencing pain sensations. In addition, you have the social context. Hence, what people mean when they refer to taste is actually what we would refer to as flavor, which is a combination of all these components.



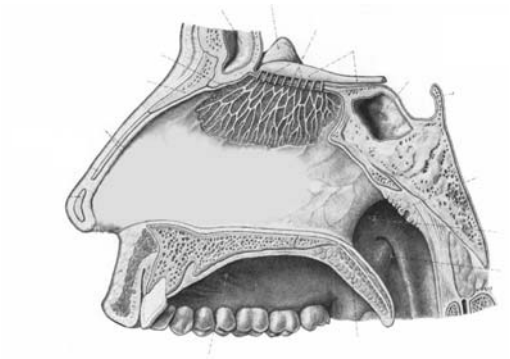


Just to give you a brief demonstration and to raise your sugar level, I would like you to take one of the candies in front of you. I would like you to pinch your nose and then put the candy in your mouth. Chew on it for a couple of seconds and then release your fingers from your nose. What you feel now is the difference between taste and flavor, which is a combination of taste and smell. When you pinched your nose, you probably didn't feel much more than a sweet taste and when you released the pinch, you felt the total flavor.

This demonstrates the fact that 90% of flavor is driven by olfaction. Interestingly, if you ask your ENT physician how many patients actually complain to them about the losing their sense of smell, he'll probably tell you that practically no one does. He will probably also tell you that the patients he does see tell him 'I can't taste my food any longer' and then go on to say that they have lost their sense of taste. However, what has actually happened is that they've lost their sense of smell.

UPPSALA UNIVERSITET

Olfactory anatomy

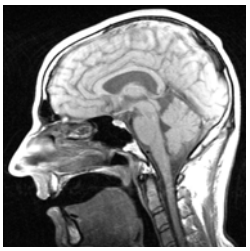


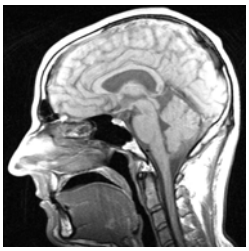
johan.lundstrom@psyk.uu.se

So why is this? Here you see the general layout of the olfaction anatomy.

UPPSALA UNIVERSITET

Retro-nasal pathway





johan.lundstrom@psyk.uu.se

As you see here, you have a soft palate in the back, that is called the velopharyngeal flap, and when you take something in your mouth and chew on it, or drink, you will see that the passage to the nose is closed. However, when you swallow, you will see that the passage opens up, allowing a passage to the olfactory receptors at the roof of the nasal cavity. This type of olfactory experience is called retronasal olfaction; a percept, that we now know, the brain treats differently than the normal orthonasal experience (smelling through the nose). Okay, now, let us talk about olfaction first.

UPPSALA UNIVERSITET

Why test olfaction?

Cognitive diseases

- Major depression
  - Patients with major depression have higher thresholds, but no difference in intensity judgments.
- Aging anorexia
  - Olfactory thresholds are higher (i.e. less flavor) which leads to a decrease in food intake and hence deteriorating health.

johan.lundstrom@psyk.uu.se

So people ask, why should we test olfaction? Well, olfaction testing is actually quite a young field, compared to vision or touch or other senses. There are papers out there that show that patients with major depression actually have a worse sense of smell than normal subjects. However, strangely, behave no differently than normal subjects when it comes to judging the intensity of an odor. There also is an increasing amount of evidence that aging anorexia, people losing

weight when they get older, may be due to decreased olfactory ability. The situation is worsened because out of fear of being sued, most retirement homes don't put any spices in the food. As a consequence they make the food taste like baby food. Everyday they're eating foods that taste like porridge, and because of their decreased sense of smell the only sensory information they receive about what they are eating is texture and basic taste but no flavor, which in turn means they eat less.

UPPSALA UNIVERSITET

Why test olfaction?

Neurodegenerative diseases

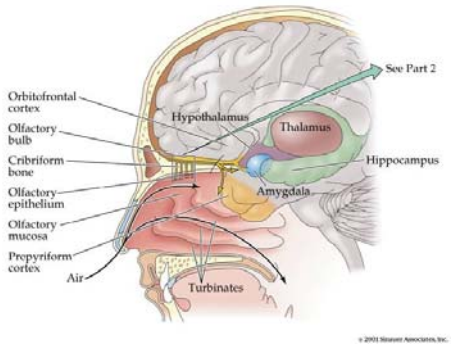
- Alzheimer's disease (AD)
  - Deteriation in olfactory thresholds, identification and olfactory memory
  - At least five years prior clinical diagnosis (90%)
  - Best diagnostic tool at the moment
- Parkinson's disease (PD)
  - Deteriation in olfactory thresholds and identification, but to a lesser degree, olfactory memory
  - Incidence of smell loss is greater then incidence of tremor
  - Even a three-item odor ID test is superior to the Mini-Mental State Examination (MMSE) in discriminating between PD and essential tremor
  - Only relatives of PD patients with olfactory dysfunction develop PD
- Multiple Sclerosis (MS)

 johan.lundstrom@psyk.uu.se

There is also huge amount of literature linking neurodegenerative diseases and olfaction. The two most common ones are Alzheimer's and Parkinson's disease. We will start by looking at Alzheimer's disease. We see a large deterioration in Alzheimer patients' olfactory threshold and olfactory memory. This marked drop can be noted in some cases up to five years before their clinical diagnosis of Alzheimer's disease. If you have some prior baseline measurements, then individual olfactory threshold is the best-known diagnostic tool we have right now to correctly diagnose Alzheimer's disease. The problem is we need to have some sort of individual baseline so we can see when the dip comes. You also have a huge literature linking Parkinson's disease and olfactory dysfunctions. We see a similarity to that pattern with regards to olfactory memory in Alzheimer's and Parkinson's patients, only to a lesser degree in the latter. Further, in opposition to the Alzheimer's patients, Parkinson's patients keep their olfactory memory performance on a high level throughout the whole disease. Something interesting is that the incidence of smell loss in Parkinson's patients is actually greater than the incidence of tremors. Studies have shown that even a short, five minute olfactory test is vastly superior to the widely used mini-mental state examination. There is also some evidence that multiple-sclerosis patients have a worse sense of smell than control subjects. The question is then why is olfaction so tightly linked to those diseases?

UPPSALA UNIVERSITET

Olfactory anatomy

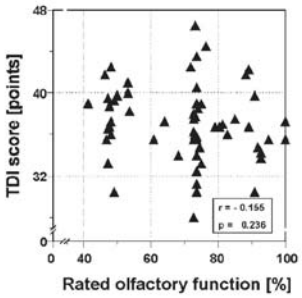


 johan.lundstrom@psyk.uu.se

If you look at the projections to the brain from the olfactory nerve one of the primary projections is to the pre-piriform cortex, which is located in the junction between the orbitofrontal and temporal lobe. You also have a major projection to the orbitofrontal cortex and the hippocampus. Something that is rarely cited is that the actual olfactory nerve endings, the olfactory epithelium, are only one synapse away from the amygdala.

UPPSALA UNIVERSITET

Validity of self-report



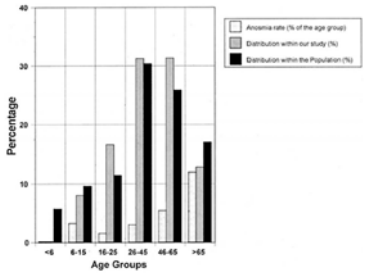
- A bad correlation between self-report and actual performance
- A clear olfactory deterioration 12 months after radiotherapy was not noticed by the patients

johan.lundstrom@psyk.uu.se

Something I am often asked is "Why can't we just ask if they have a normal sense of smell?" This is what is normally done in studies of our other senses. However, the correlation between actual function and perceived function is very bad. This is a combination of threshold discrimination and identification scores and how they rate their sense of smell. As you can see, there's virtually zero correlation there.

UPPSALA UNIVERSITET

Validity of self-report



- 77% of elderly with smell loss reported normal sense of smell
- Of the elderly reporting normal sense of smell, 13% had total anosmia

johan.lundstrom@psyk.uu.se

In a study where olfactory sensitivity was measured, 77% of the elderly with a major loss of smell reported that they had a normal sense of smell. Moreover, in a recent study, out of the elderly that actually reported 'I have a normal sense of smell,' 13 of these had no sense of smell whatsoever. Compare this to visual research where you never will see that out of the group that says 'Yes, I have 20/20 vision,' when in fact they are actually totally blind. This would just not happen.

Unidentified Speaker: Every single sensory function that you look at, people overestimate their real ability, the hearing, the visual, I'm sure it's exactly the same.

Lundstrom: Right, but the difference is here they have no function whatsoever.

Unidentified Speaker: Indeed, when blind people claim that they have perfect vision, it would be a symptom of psychosis.

UPPSALA UNIVERSITET

What is measured?

- Identification
- Absolute threshold
- Discrimination
- Hedonic preferences

johan.lundstrom@psyk.uu.se

Lundstrom: When we talk about olfactory measures, we usually talk about these four tasks. I would like to only talk about identification and threshold due to time constrains. Moreover, I will focus on the more "quick and dirty" ways to measure this. If you want to do it properly, as with all the senses, you have to take them into the lab and spend around twenty minutes on each test. However, I'm going to focus on how to do this in the field and in a sound way. I'll start by talking about the identification test.

Lundstrom: Common olfactory tests, especially here in the states, are those scratch-and-sniff tests. I brought some with me. These come in a variety of forms and are easy to store and to administer. However, the problem with these is that some of them are "do-it-yourself tests," which means that some people scratch more, some people scratch less. In fact, there's a study demonstrating that some of the olfactory effect I talked about previously in Parkinson's was due to that Parkinson's patients did the scratching themselves. People with the worst sense of smell scratched only once and people with the best sense of smell scratched more than once. Which is a problem: you don't have control over this important factor.

UPPSALA UNIVERSITET

Clinical tests of olfaction

- Identification tests
  - UPSIT
    - ~ 20 minutes (40 item)
    - ~ 6 minutes (10 items)
  - Sniffin' Sticks
    - ~ 8 minutes (16 item)
    - ~ 5 minutes (12 item)

johan.lundstrom@psyk.uu.se

So the one I favor right now is the Sniffin' Sticks. This is a newly-developed method - about four years old. It's a highly validated method and easy to administer and to transport. I brought some with me but you might want to avoid sniffing pen number four, that is the fish odor. These are normal felt-tip pens filled with an odor. You can fill these with whatever odor you want to have in there, as long as they are diluted in some sort of liquid. Moreover, they last a long time – we have pens that we filled up five years ago, and they still have a strong odor. Right now, by using pictures instead of words they are trying to standardize these tests so they can even be used in children as young as three years of age. Three types of tests can be bought: identification, discrimination, and threshold tests.

UPPSALA UNIVERSITET

Clinical tests of olfaction

- Sensitivity (threshold) tests
  - Sniffin' Sticks standard
    - ~ 10 minutes
    (~6 with new method)
  - Intensity estimates as base for threshold calculation
    - ~ 3 minutes
    - Less accurate

johan.lundstrom@psyk.uu.se

Now, we turn our attention to threshold testing. Threshold testing is the most complicated test. However, they have a standardized set that covers more or less the whole range of people's olfactory capabilities. It takes around ten to twelve minutes to administer. A couple of weeks ago, a new paper came out where they tried a new method. They now claim that you can get an accurate threshold measure in only six to seven minutes per subject. Moreover, they showed a really good correlation between the older method and the newer method. But what we're trying to do in this NSHAP study is a new thing. We're actually trying to estimate the threshold by using intensity judgments. We think we can get down to around two to three minutes per person by using around five different concentrations.

UPPSALA UNIVERSITET

Taste

johan.lundstrom@psyk.uu.se

Okay, that is all I have to say about olfaction. Let's go to taste. Why measure taste?

Well, among other things, we know that several studies have shown that obesity is tightly linked to taste performance. Moreover, something I really want to stress is that if you really want to go into taste research, you can sit down and read most articles there is on taste during one week. Very little research has been done to date on taste perception, thus leaving the field wide open for new and novel research ideas. The most common type of research in the field of gustation asks subjects for descriptors of food stimuli. However, this is not a measure of taste abilities. These studies are actually measuring flavor or interaction between the sense of taste, smell and trigeminal system (pain).

UPPSALA UNIVERSITET

Why test gustatory functions?

- Obesity
  - Taste sensitivity is connected to obesity which will soon be the number one death cause
- Relativly un-explored sense
  - Few studies have looked at connections between taste functions and health due to difficult testing procedures and lack of awareness

johan.lundstrom@psyk.uu.se

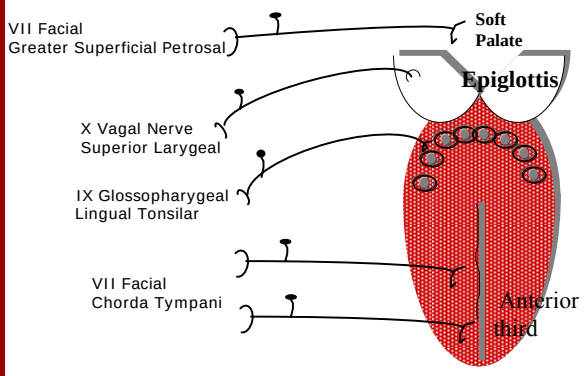
What to test?

- Identification
- Absolute threshold
- Discrimination
- Hedonic preferences

johan.lundstrom@psyk.uu.se

In this case, I will talk about identification, because I think that measuring threshold and discrimination is a too difficult task to perform in the field. One of the problems with taste is that you have to decide if you want whole-mouth testing or regional testing. The reason behind that is if you do regional testing, you're going to activate different nerves.

Anatomy of Gustatory System



johan.lundstrom@psyk.uu.se

There are three general nerves connected to the general taste sensation in the mouth.

Taste

Whole mouth testing or regional testing?

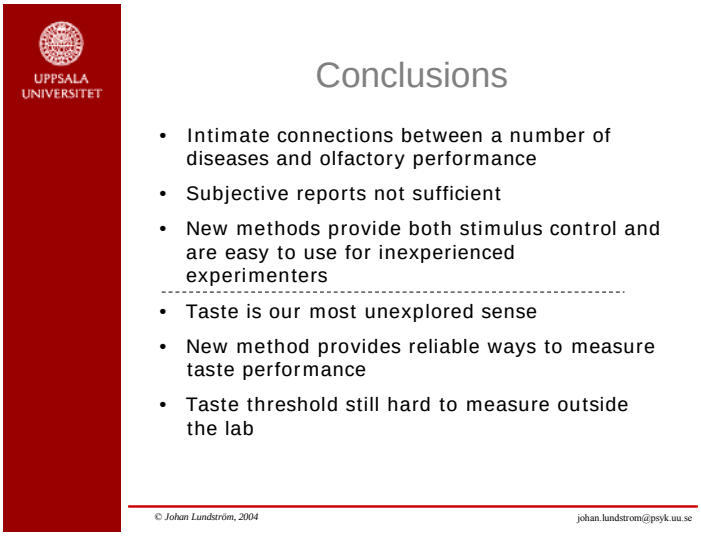
- Taste strips
  - Fixed amount of stimuli
  - Easy to use
  - Easy to bring
  - Easy to train
  - Can be used for both regional and whole mouth testing
- Taste spray
  - Liquid in spray bottles
  - Varied amount of stimuli
  - Only whole mouth testing

johan.lundstrom@psyk.uu.se

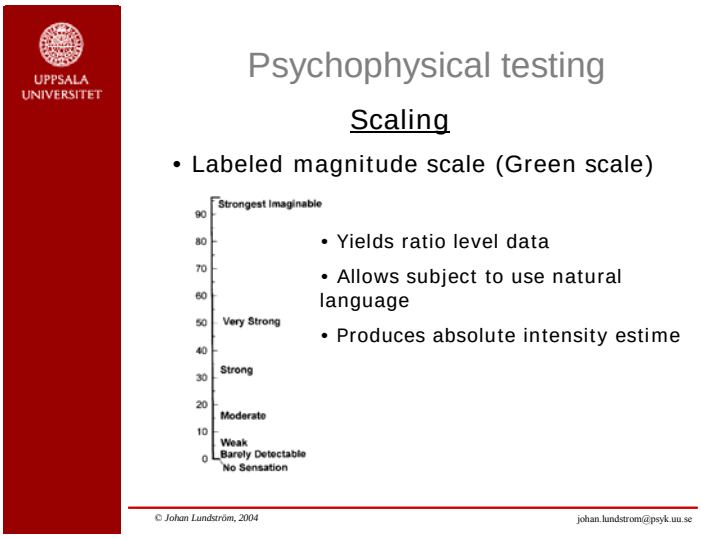
There are a few methods out there. The one that I favor is the taste strips. They're small filter papers that have been dipped in a solution, which gives away a more or less equal amount of stimuli in each simulation. The good part about these is that you can decide if you want to perform whole mouth or regional testing. You can do whole mouth by actually dipping the strip on different parts of the tongue

Unidentified Speaker: Is that something you do in the instructions when you give it to the participant? Do you say, you know, you swish this around in your mouth?

Lundstrom: They should not do it themselves. The experimenter should do it so they have a standardized procedure. However, the good thing about these is that it is easy learning how to administer them correctly. An alternative to these strips is taste spray. They're small bottles that spray the mouth with a liquid. However, the problem with this method is that you get, depending on how good the spray mechanism is, a considerable variation in the amount of stimuli you administer, and only whole mouth testing can be performed. Other problems with this method are that you need to have some sort of water with you to clear the mouth after each stimulus. That is not that essential with those taste strips.

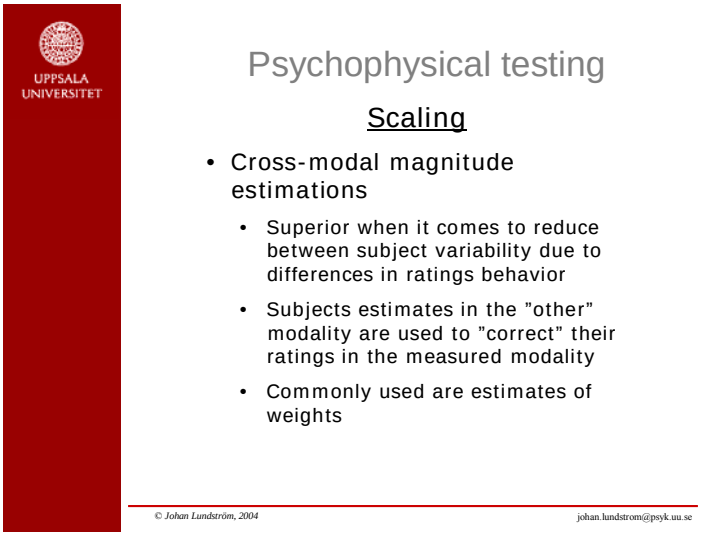


Conclusions. When it comes to olfaction, we have an intimate connection between several diseases and olfactory performance, due to the close neural connections between olfactory centers in the brain and the neural substrate of the diseases. Moreover, we now know that subjective reports are not sufficient. We have new methods for measuring olfactory functions, the sniffing sticks. These provide both good stimulus control and are easy to use for inexperienced experimenters. It takes around five minutes to teach someone to administer them.



Taste. The good thing about taste is that it is a relatively unexplored sense, which is also its great disadvantage. So if you do something in taste research, you're pretty sure that you're going to be the first one. We also have those new methods to measure taste functions, these taste strips. They are easy to transport since you can bring them with you in your pocket.

Now, let us talk about psychophysical scaling. I don't know how much you know about this, but the psychophysical field started to turn towards label magnitude scales a few years ago. These scales provide much better data than another scale that has been used to date.



In this instance, you get data on ratio level. Data on this level are more suitable for advanced statistical calculations. You also allow the subject to use their natural language. Instead of using abstract numbers such as 0 to 10, they can tell you in a verbal language how intense something is. Moreover, if you have time, I would really recommend the use of crossmodal magnitude estimation. What is crossmodal magnitude estimation? Let's say you want to measure touch. We know that people have different ways to communicate sensory experience, or any experience for that matter. Moreover, we know that women are more prone to use the middle-end of the scale, and that guys are more prone to use extremes. So, we are going to have a huge variation between subjects in how they use scales. By using ratings on an independent scale, in this case we might ask them to estimate a couple of weights, and use that as a baseline for their ratings behavior. This rating behavior can then serve as a baseline from which we can adjust every behavior measure we are collecting according to how they normally rate things.

McDade: How reliable is a single measure? So if you measured someone in the morning and the evening, would you get the same results? Before and after a meal? Or for women, over a month, in terms of their stage in the menstrual cycle?

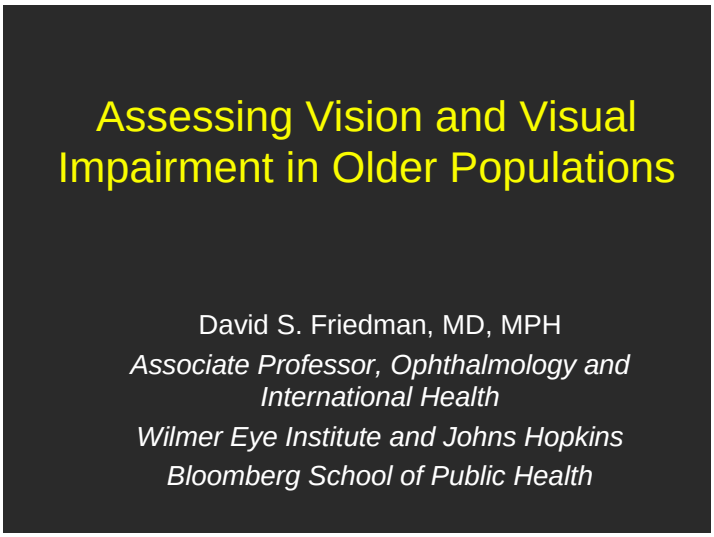
Lundstrom: The problem is that, and I don't know how accurate this statement is in taste, but in olfaction, you have a higher variation within a subject than between subjects. This means that if you measure an individual ten times, you can have a huge variation between those ten times. Moreover, over the course of a menstrual cycle, we know that women have their highest sensitivity just before they ovulate. So indeed, it seems like our hormones can regulate our sensitivity.

McDade: So if you were doing survey research and you only had one time to measure an individual, is that one worth doing, given that there's so much variation?

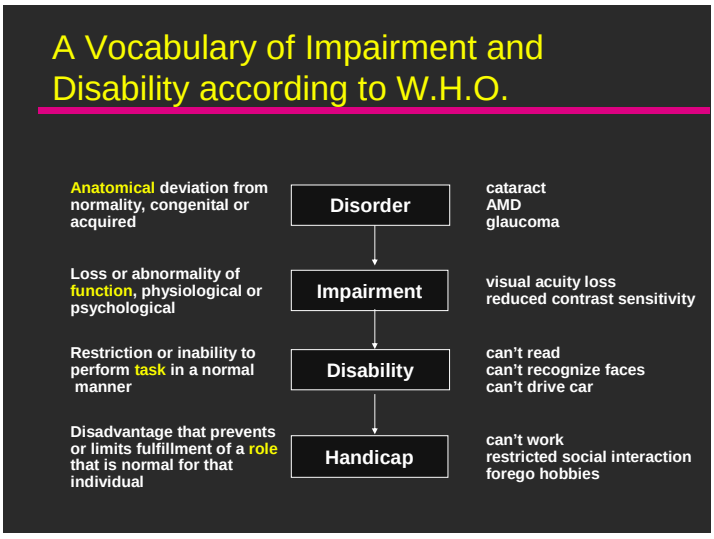
Lundstrom: It depends on what we are measuring. I talked about absolute threshold, i.e. their sensitivity. But if you talk about identification performance you will not find these strong variations due to the high intensity of the stimuli used.

Assessing Vision and Visual Impairment in Older Populations

Speaker: David S. Friedman

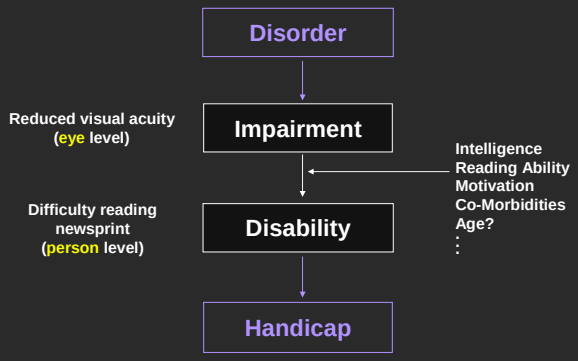


I'm going to talk about visual impairment. There's been a lot of research at Hopkins on this topic. I've participated in a bunch of it, and some of it was done by others, particularly Gary and Sheila, who've been incredibly productive in this area.



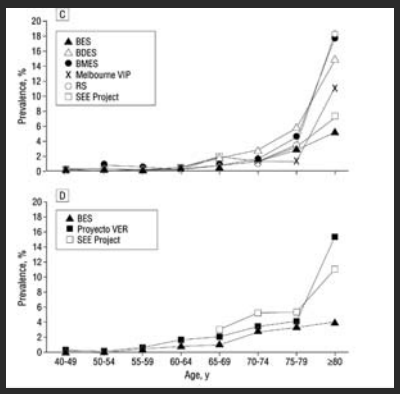
It's nothing new to anybody here: thinking about how we think about impairment, disability, and handicap. I'm sure you've thought about all this as well, but I have some eye examples, and I think it's a good way to think through it. An anatomical deviation would be a disorder, and so people get cataracts, they get macular degeneration, which takes away the central vision, they get glaucoma, which takes away side vision and has specific impairments that occur. Cataracts might cause reduced vision for driving and reading, but macular degeneration takes away your total ability to read, and there you get a disability. Ultimately, you become handicapped and unable to perform certain things that you'd like to do.

An Example



One specific example might be reduced visual acuity as the impairment. Then you have difficulty reading newsprint, which is actually an incredibly frequent complaint from reduced visual acuity, and one of the reasons that people come in, ultimately, for treatment and care. Also it's influenced, and we've talked about this, by all these other things: the cognitive issues, your native intelligence, your ability to read to begin with, motivation, and co-morbidities – if you have a shaky hand, that might influence your ability to use what limited vision you have because you might not be able to hold the reading material comfortably at a proper distance.

Prevalence of Vision < 20/40 in Population-Based Studies



Whites

Blacks and Hispanics

We've looked now at population-based studies of visual impairment, and this just gives you a sense of how incredibly rapidly visual impairment increases with age. As you get into this 70-74 age range, rates are still relatively low in both of these groups, but then there's this exponential increase as you get into the 75 and older group. This is vision less than 20/40, and the reason that cutoff has been chosen is it's been shown that there are limitations, functional limitations at that cutoff, and they are important and meaningful ones.

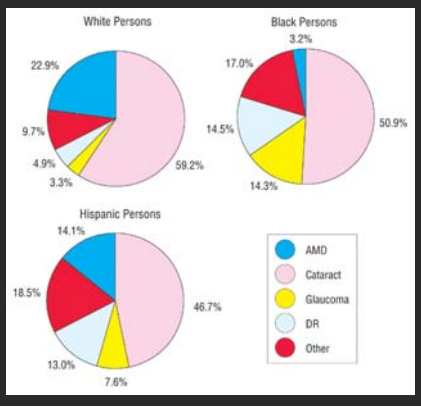
I've just studied glaucoma in a very elderly population, and there's also an amazing take-off in prevalence among older Americans.

The Impact of Low Vision Increases with Age

- Fewer than 20% of people under the age of 45 with low vision are severely limited in their ability to work or keep house
- More than 40% of people over the age of 65 with low vision report significant limitation with daily activities

Low vision impacts people differently at different points in life. When you're younger, you're more able to adapt, because you have other reserves, as people were talking about previously: you have the ability to rely on your strength to help with your mobility, to rely on other factors to evolve, maybe, to accept this low vision. Whereas when you're older, you really become limited in your activities of daily living when you lose your visual acuity. I want now to give you a sense of the causes of visual acuity loss.

Causes of Vision < 20/40 in Population-Based Studies



Macular degeneration is, among whites, a major cause. Cataracts now are likely to be treated at an earlier state, but still many people wait to present for care until they're substantially disabled by this condition. Different diseases affect different races, so African-Americans are much more likely to have glaucoma, and with Hispanics it's sort of in between those two in terms of the diseases. Different diseases affect different individuals, and even though we talk about visual impairment, these diseases impair vision in dissimilar ways.

So even though one researcher, Salive, looked at people in their own homes, what we're going to measure in this study is habitual vision. We're not going to measure best corrected vision. Everything I've shown you is best corrected vision, so when we talk about visual impairment, we're talking about impairment after caretakers tried to refract and put them in glasses – they still had residual vision loss. When we go into the homes, we're going to see even higher rates, because individuals who have their current glasses that are not recent are going to have visual disability as well.

We'll see much higher rates, and you can see here, 28-33%, if you sum, of the 20/40 and worse vision. The associations with visual loss and functional loss have been shown in multiple studies. I just want to give you a sense of some of the functional loss, and then I'll talk about the methods of vision testing.

Visual Impairment at Home

- VA < 20/200 = 4.5%
- VA < 20/60 – 20/200 = 14.5%
- VA < 20/40 – 20/60 = 14.0%

Mean age = 79

Previous Research

Salive found that individuals with decreased vision examined in the home had:

- Greater difficulty with ADLs
- Poorer physical strength
- Greater likelihood of new onset physical disability developing

For individuals with visual impairment it has been shown that they have greater difficulty with ADL's. This was just testing the vision in the home, one time, and looking at what subjects reported. What was really interesting was that subjects had a greater likelihood of new onset disability, they had greater likelihood of becoming physically weaker and less mobile when you came back a year to two years later. Just on the basis of vision. If you just looked at their visual acuity at baseline they were less likely to become so. It kind of leads into a story: you start getting visually impaired, you become more afraid to move around, you become weaker, and you end up being more likely to decline. I really think that is the mechanism through which it works, and I'll give you some evidence for that.

This is the Salisbury Eye Evaluation Project, which was a longitudinal study conducted on the Eastern shore of Maryland, and it's had four rounds over a decade. This study demonstrated that people with visual impairment have greater social isolation, they don't participate in social activities, they avoid religious activities, and they have difficulty with ADL's and IADL's.

Previous Research

- SEE study found VA < 20/40 resulted in decreased function and greater isolation
  - OR = 1.7 for no social activities versus any
  - OR = 2 for religious activities
  - OR = 1.8 for any difficulty with ADL
  - OR = 2.45 for IADL difficulty

Auckland Hip Fracture Study

- Binocular VA < 10/60 associated with hip fracture over 2.5 year period (OR = 1.5)
- VA < 20/100 OR = 2.4
- No depth perception OR = 6.0
- Self-reported poor vision OR = 1.4
- 40% population attributable risk

If you look at this whole issue of mobility in terms of hip fractures, it has been reported that individuals had a much higher likelihood of hip fracture if they had poor vision in one eye. And if an individual lost depth perception -- and this is one of the few studies where that's been shown -- he had a tremendously high risk. The authors felt that vision alone had a very high attributable risk for hip fracture.

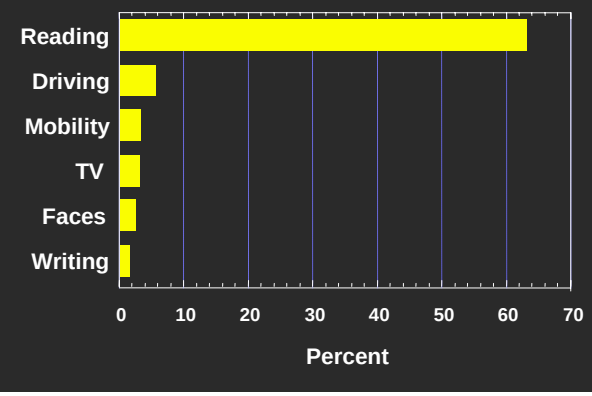
Beaver Dam

- Used 20/20 versus 20/25 or worse
- Visions were obtained five years previously
- Results reported for those over 60
- Almost three times as likely to report falls in past year (12% versus 4%)
- Hip fractures more common (5.2% versus 1.4%)
- Walking speed was slower

This was found as well in another large population-based eye disease study in Beaver Dam, Wisconsin, where they found a much higher rate, even at this subtle difference in vision. This is people who had early cataracts, a lot of them, and subtle worsening of vision from 20/20, yet they had a much higher rate of falls, so it may be that the loss of vision, that is, the change from good vision to worse vision, that is an important contributing factor.

Well, what do people complain about? They mainly complain about reading. If you administer a self-report, you can see this is the biggest complaint of low-vision patients. Many of them stopped driving, so a number of them complain about it, but then they don't really notice it. There's an inability to recognize that you're limited by your vision. Many patients that I've taken care of, just on a clinical level, don't recognize their lost vision, and I think that's a major problem with slow sensory loss. You slowly avoid activities that you didn't realize you weren't doing. So if you have a shoulder ache, you don't reach up. You don't notice that you never reach your right arm above your head, and I think the same thing happens with vision-related activities.

Chief Complaints of Low Vision Patients



Visual Acuity - ETDRS Chart

- Standardized illumination
- Geometric size progression
- Uniform spacing
- Letter-by-letter scoring
- Forced-choice testing

Let me get into the methods of vision testing. What I would have liked to have done, had we had more time in this upcoming study, is to illuminate the letter chart with a self-illuminated screen that would provide equal illumination across the screen. You can see that there's some nice vision testing instruments, and these have been around for about 30 years. The letters are equally spaced, and they come down in a proportionate fashion. It's done in such a way that these are on a log scale. So it's a log-decrement, of the minimal angle of resolution. So it's called a log-MAR scale. So that allows you to do statistics on these, treat these as linear variables. You count the number of letters, and you know where they end. They have also picked letters that have similar shapes. So a 'D' and an 'O' are okay, and 'V' and 'K', but if

you were to put a 'Q' up there, it would be harder to recognize. So they actually went through and looked at recognition response characteristics. One of the key elements you have to do in any vision testing, besides standard illumination, is a forced choice testing, because people say 'Oh, I don't know,' and even though they have only a 1 out of 26 chance of guessing the letter, they manage to get five correct in a row. They're seeing it, they just don't believe it. There's different 20/20's. There's 20/20 where it's like boom boom boom boom boom, and there's 20/20's where it's like…[long pause]…they get it, but you have to push them.

I'm going to talk about other things that we're not going to do in this study, but there are other ways of looking at vision that are very important that I think you'll need to be aware of if you're thinking about vision in your patients or studies. 'Acuity' describes the eye's ability to discern fine detail at high contrasts. That's really what we're doing, but that's not all there is to vision. Acuity can remain intact when other parts of vision decline, and they're not sensitive to problems in contrast or loss of side vision. So people have developed these other tests.

Most studies define visual impairment by acuity loss

- Acuity describes the eye's ability to resolve fine detail at high contrast
- Acuity may remain intact when other aspects of vision decline
- Acuity tests are not sensitive to problems in seeing large low-contrast objects (e.g. faces)
- Acuity can miss loss of side vision

Contrast Sensitivity
Pelli Robson Letter Chart

- Large letters
  - Unaffected by acuity
  - Insensitive to refractive error
- Familiar, quick, and reliable

This is a uniform size, but contrast goes down throughout the chart. That's a contrast sensitivity test.

You can do the same thing with glare -- you can add glare to the test, because certain cataracts and other disease conditions can be affected by glare. No study that I know of has found much in the associations of glare disability and overall function or relationships. You may also lose depth perception, and there are 3-D tests.

Glare Sensitivity - Brightness Acuity Tester

- Can be used with any vision test
- Full-field glare
- Reliable and sensitive

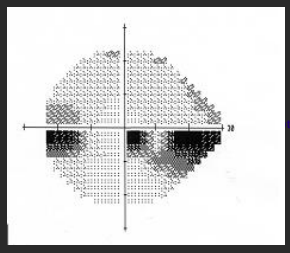
Stereoacuity - Randot Circles

- Familiar clinical test
- Reliable and valid for children
- Not widely used in adult populations

This is called a Randot test, and so that these stick out, you can see them coming out of the page with these polarized lenses, but there's not much data on the use of this test in older people.

And then there's glaucoma. Here I am showing a moderate defect. This would clearly interfere with reading. This is the center of vision and just below it you've got a dense blind spot. You could lose everything -- this could be black for 360 degrees and you'd still have perfect central vision. So you can have 20/20 vision with end-stage glaucoma. If you walk out there and you just test vision, you're going to miss this constriction of the visual field in some people. Fortunately, that's a relatively small proportion of visual impairment, but you can be legally blind from field loss.

Visual Field - Humphrey Automated Perimeter



- Visual field loss is rarely identified by visual acuity testing
- Mobility is decreased with field loss
- Legal blindness can exist with 20/20 vision

Self-Reported Visual Function

- National Eye Institute-Visual Function Questionnaire (NEI-VFQ)
- Activities of Daily Vision Scale (ADVS)
- Visual Function 14 (VF-14)
- Others

Self-Reported Visual Function Questionnaires

- Strong association with visual acuity, but not completely determined by VA
- Internally consistent
- Associated with function
- Improvements shown in patients undergoing cataract surgery (responsive)

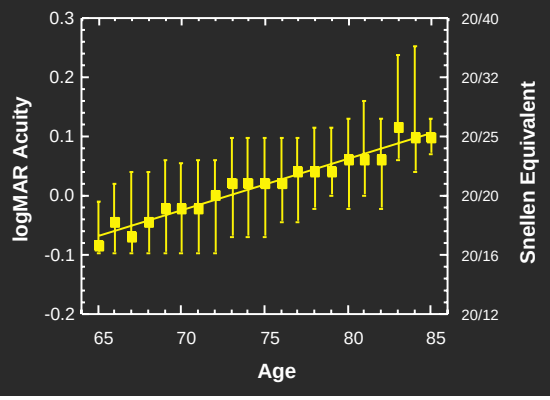
There are self-reported visual function questionnaires, so there's the NEI-VFQ, the activity of daily vision scale, the VF-14. These have all been validated, they have all been shown to have strong associations with visual acuity, but they are not completely determined by visual acuity. They're internally consistent questionnaires, they're associated with function, and they show responsiveness. So if you take out people's cataracts, they do much better on these questionnaires, and that's been shown in multiple studies. They really are capturing visual function, and they are somewhat independent of central visual acuity.

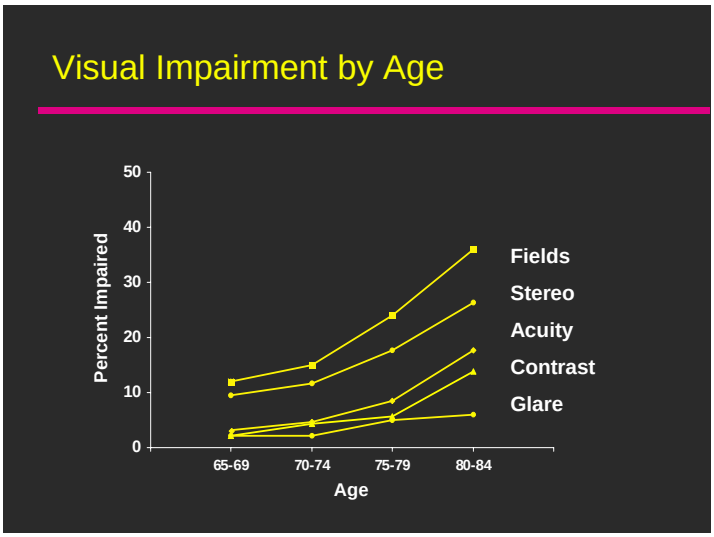
Salisbury Eye Evaluation Project (SEE)

- Population-based study of the impact of eye disease and visual impairment on physical disability and quality of life
- 2520 randomly selected residents of Salisbury, Maryland, between the ages of 65 -85
- 20% of sample African American
- Longitudinal study

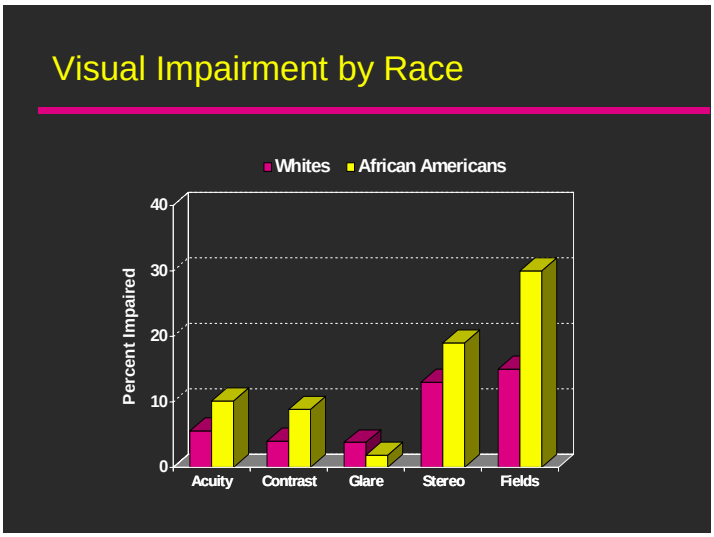
I mentioned the Salisbury Eye Evaluation Project. It enrolled older individuals and tried to oversample African-Americans. I'm just going to show you some associations with acuity and age, so you can see that acuity declines with age, which goes along with the fact that I showed you visual impairment and age-related eye disease are both occurring.

Visual Acuity by Age

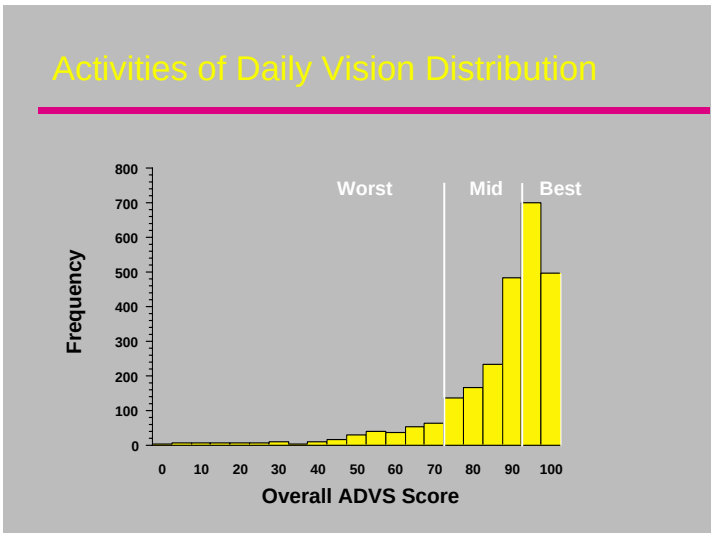




If you look at fields, stereo, and contrast -- glare seems somewhat flat, but with all the other ones -- there's this big uptick at 75 to 79, and it's all related to cataract, macular degeneration, and glaucoma, really.



There is a higher prevalence of disease among the African-American community than among the white community. That's been shown in several studies, and I think it's believable. It seems they have more cataracts and that maybe there's less likelihood of undergoing cataract surgery. There are some clinical trials where people were followed and developed cataracts, and the African-American cohort did not go for surgery as often as the white participants.

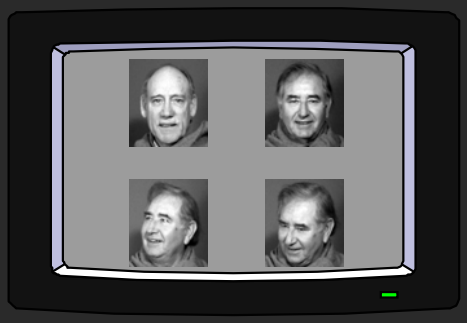


Odds Ratios for Overall ADVS

Demographic Variables		Vision Variables	
Age (per decade)	1.04 (0.75,1.46)	Acuity	2.39 (1.77,3.23)
Female Gender	1.76 (1.47,2.11)	Contrast	1.85 (1.35,2.55)
African American Race	1.12 (0.91,1.37)	Glare	1.84 (1.28,2.55)
MMSE (per point)	1.06 (0.99,1.14)	Stereo	1.34 (1.22,1.48)
Education (per year)	1.04 (0.99,1.11)	Fields	1.37 (1.19,1.58)
Other Diseases	1.36 (1.07,1.51)		
Depression	1.27 (1.07,1.52)		

This just shows that you can get a distribution on one of these questionnaires, and if you look from the lower tertile to the upper tertile, you can see that it's associated with the vision variables, but it also has other associations as well, these vision scales.

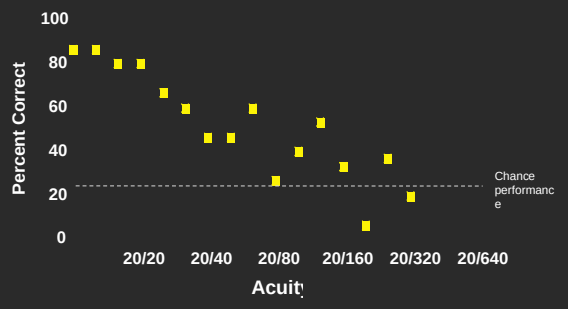
Face Recognition Test



Which picture is of a different person?

You can do other tests of function associated to recognition. Face recognition with visual acuity: obviously, it's going to decline. These are not age-adjusted. While this line appears almost perfectly straight, it is not as well-correlated when you actually adjust for age.

Face Recognition as a Function of Visual Acuity



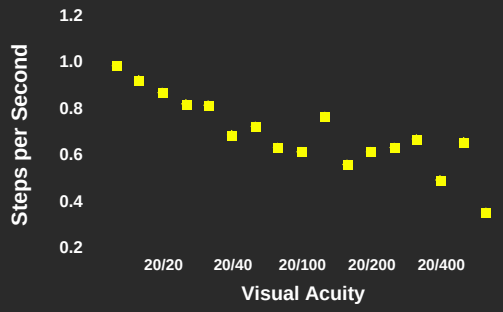
Visually impaired participants recognize about 15% fewer faces per line of acuity loss

Stair Descent

You can look at how people go up and down stairs, and again, this is not purely age-adjusted. Clearly visual acuity is associated with how fast people go up and down the steps.

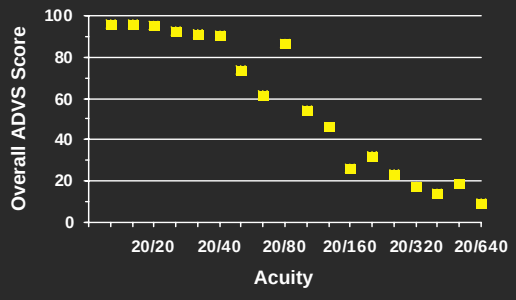
You can look at self-reported difficulty with visual tasks, and they're highly correlated with visual acuities and with contrast sensitivity.

Association of Stair Climb Speed with Visual Acuity



Visually impaired participants walk down stairs about 7% slower per line of acuity loss

Self-Reported Difficulty with Visual Tasks



Visually impaired participants report about 10% more difficulty per line of acuity loss

Conclusions: Vision Impairment and Disability

- Prevalence of visual impairment increases with age for multiple vision measures
- Vision loss is associated with depression, social isolation, and loss of independence
- Difficulty with daily activities is determined by multiple components of visual function; not just acuity
- While other measures of vision are important, visual acuity is the strongest predictor of functional impairment as measured both physically and by patient report

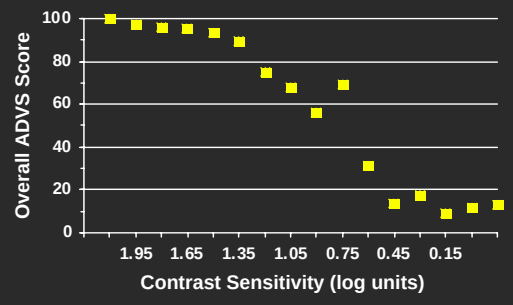
So, in conclusion, prevalence of visual impairment increases with age for multiple measures of vision and visual function. Even if we measure it by visual field, contrast sensitivity, or glare testing, every test shows declines. Vision loss has associations with depression, social isolation, loss of independence, and I imagine it will with self-image as well. Difficulty with activities of daily living is determined by multiple components of visual function, not just acuity. But, if you have one measure that you get three minutes to do, I think a high-contrast exam of vision is the best. It's the strongest predictor in all these studies. So that's it.

Waite: We talked a little bit earlier about how an impaired sense of smell affected women's sexual desire, but could you argue maybe, especially for men, that impaired vision might improve it?

Unidentified speaker: With regard to depth perception, what would it require and what would it cost to do a measurement of depth perception in the context of a population-based study?

Friedman: Well, the problem with this – at least the one thing that's been done, the stereo test – and I actually talked to the statistician who's really excellent on that study, and she said that there was just so much impairment on that test that you ended up having the whole thing shifted up, so it was very hard to identify depth as influencing any other factors. So one would have to develop something new. All this stuff's been developed for children, because children get Amblyopia, or 'lazy eye,' they can't stereo, and it's all geared around that. It's not really geared around functioning with depth. They have a 'put your key in a lock' test which would involve some hand coordination as well as depth perception -- some of that stuff is decent, but I think you'd have a hard time testing it with a validated instrument.

Self-Reported Difficulty with Visual Tasks



Visually impaired participants report about 15% more difficulty per line of contrast loss

Pierret: I was just wondering how difficult it is to get untrained people, or semi-trained people, out to do the eye test. Do you have to be a certain distance away and all that?

Friedman: Yes, it's set at a fixed distance. You need a piece of tape or something that you're just going to lay out, and then if the subject can't see it, the subject has to come a little closer and this has to be recorded. The biggest issue is to make sure that the technician is encouraging the person to guess, because if the subject just stops trying, that is a real problem. So before we do it in this study, I'll come out and train and make sure, and then I can periodically come out and view what they're doing or something along those lines. We have a lot of different people who've been trained to do it.

Pierret: Presumably, some of these things are easier to correct than others.

Friedman: Yes.

Pierret: So visual acuity, I would guess…

Friedman: Well, refractive error and cataract are the two most common causes. I didn't show you habitual visual acuity, but refractive error is about sixty percent of decreased vision. So if you looked at low vision in older populations, a lot of them just need glasses, and that's easy, but-

Pierret: But studies that show they're associated with hip fractures are corrected 20/100.

Friedman: Well, the Auckland, I believe, was uncorrected, the Beaver Dam was best corrected. But the Salive study that showed real declines in physical function, were uncorrected. But cataract, obviously, it's a fun surgery to do. It makes people see, and can take as little as ten minutes to perform.

Brauner: So evaluating visual acuity in people with dementia…

Friedman: It's a huge challenge. I just completed a study of 28 nursing homes and I examined every person with decreased vision. There are actually charts developed for testing pre-verbal children called preferential looking charts, and an observer looks through a peephole while the person being tested is shown a target with a grated line on it. The grating is finer and finer on subsequent cards, and eventually it is very hard to see. The observer watches to see if the subject is looking at the grates to determine the visual acuity. I didn't show you an example here because we're not going to be doing that. We did find you could improve testability with this technique. We had a mean Mini-Mental State score of 12.4 in our study population, and we were able to test 90+%. For about 15% of the study population we used this preferential looking test, but it is tough to do. It is also hard to decide whether or not it's visual impairment, or cognitive impairment and just inattention that is causing the subject to test poorly.

In one other study the authors found an association with visual fields and mobility, but they used a terrible field test, and I just think it's just people not paying attention. When people don't pay attention they can't walk well, so it's a tough one. Generally, though, for central acuity, if you keep forcing them to guess and get back on track, most people can do it unless they're severely demented.

Hearing Screening in Population-Based Research

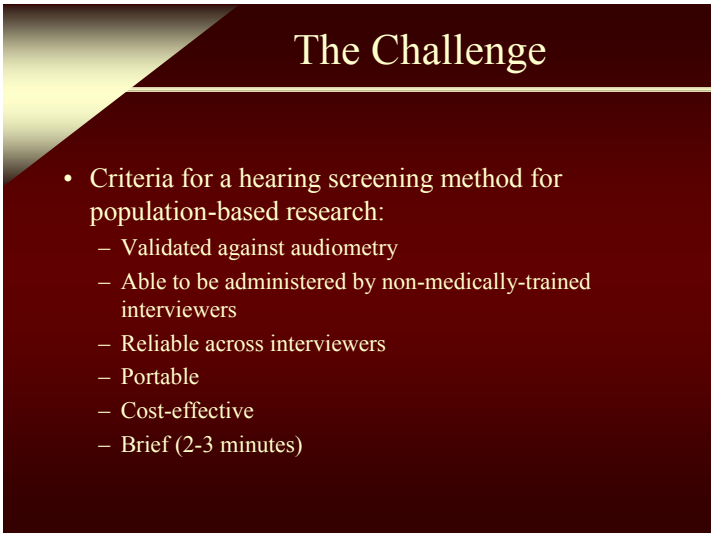
Speaker: Erin York

As a sociologist, I never imagined that I would be studying hearing loss and learning about various ways of assessing older adult hearing. However, through my work with the National Social Life, Health, and Aging Project (NSHAP), I have learned a great deal about hearing screening and am convinced of the importance of learning more about how hearing affects older adults' health and quality of life. I'm going to talk a little bit about the available methods for hearing testing, the challenges presented by a population-based study, and new technologies that are on the horizon for hearing screening in population-based research.

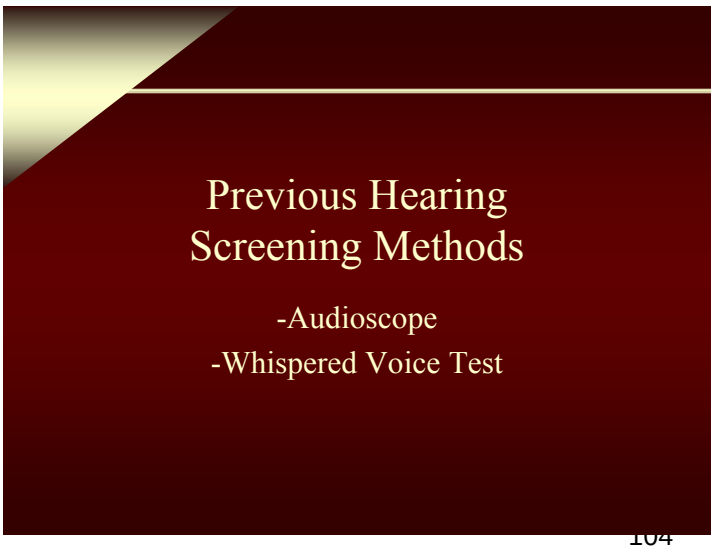
Hearing loss is often referred to as the third most prevalent chronic condition among older adult. Unfortunately, we don't have good population-based research on hearing loss among older adults.

We estimate that about 30 to 46 percent of older adults have suffered from hearing loss, and, as David Friedman showed, this goes up dramatically as you age from 45 to 80. Ninety percent of adults over 80 have hearing loss. And as Sara Leitsch mentioned at the start of this panel, hearing loss is related to several other aspects of social life and physical health and mental health, like depression, low self-esteem, social isolation, and functional decline. So, in order to both establish the prevalence of hearing loss and understand its broad ranging effects, we need to be able to study hearing loss in a population-based study where we're also getting measurements of social factors and functional and mental health.

When I started learning about hearing testing, I found out that the gold standard for measuring hearing is pure-tone audiometry. Audiometric testing requires the respondent to put on headphones, and the tester plays a series of sounds with certain thresholds, and the respondent raises his hand to indicate when he hears the tone. The tester is then able to establish the respondent's hearing thresholds and diagnose a hearing loss. However, this method requires an audiometer, which is very expensive, a sound-proof booth, and a certified audiologist. And it's also quite a long test. So this is not really something that we were able to take out into the field for a population-based study.



My challenge in working on this was trying to find some method that could be used in a population-based, interdisciplinary study such as NSHAP. We needed something that would be validated against standard audiometry, something that could be administered by non-medically-trained interviewers, and something that would be reliable across the many interviewers that we would be using in different locations. We also needed a testing method that would use equipment that is portable, since we're doing interviews in the home. Finally, we needed something that was cost-effective, since hearing testing is just one of many measurements we were hoping to take, and brief. I was told we needed to limit the administration time to two to three minutes. This was a tall order.



In looking at the literature, I initially identified two possible methods. First of all, a portable audioscope, like this one made by Welch/Allyn, seemed like a possibility.

Audioscope

- Tones delivered through probe tip inserted into auditory canal
- Sensitivity = 94%; Specificity = 69%
- Portable
- Brief
- Some training required
- Cost = $500-600

I consulted other researchers who immediately pointed me to the Welch/Allyn Audioscope as a kind of "portable audiometer." Because it is inserted into the ear canal, it controls for some environmental noise, so it is something that could be administered across a variety of locations. It has good sensitivity, it's portable, and it's brief. But, the cost for us was prohibitive. At a cost of about $500-600 per Audioscope, it would have cost us $50,000 to $100,000 just to outfit our Field Interviewers.

Another test, which is quite inexpensive, is the Whispered Voice Test. And amazingly, I found this used in many studies of hearing loss – although it has been largely limited to small, clinical studies. In this test, the subject sits in a chair, and the examiner stands behind at arm's length. The subject covers one ear with his or her index finger. The examiner whispers combinations of letters, and then scores the test based on whether the patient is able to repeat those combinations.

Whispered Voice Test

- Examiner stands at arm's length behind seated patient
- Patient covers one ear with index finger, rubbing in a circular motion
- Examiner whispers a combination of letters and numbers (e.g. 4-K-2)
- Pass = 3/6 numbers or letters correct

The instructions are very imprecise and where there are specifications, such as the distance at which the examiner stands, the specifics vary across studies.

Whispered Voice Test

- Cost-effective
- Portable
- Brief
- Sensitivity = 90-100%; Specificity = 70-87%
- Wide variation in methods
- Inter-examiner variation

The Whispered Voice Test is cost-effective and brief, and these small clinical studies have shown fairly good success with it. But obviously, there's a wide variation in the methods. Because the protocol for this is not very well-defined, it would be difficult to try to administer this consistently across a large population-based survey. Sounds in the home environment, as well as inter-examiner variation – in terms of how clearly and loudly the examiner would whisper – would likely render the results unreliable.

Self-Reported Hearing Tests

- HHIE-S
- Global Question

Self-reported hearing tests emerged as another option. The Hearing Handicap Inventory for the Elderly Screening version (HHIE-S) is a ten question index to measure social and emotional handicap from hearing loss. It includes questions like, "Does a hearing problem cause you to feel embarrassed when meeting new people? Do you have difficulty hearing family and friends when you're in a restaurant? Do you have trouble hearing people when they whisper?"

Hearing Handicap Inventory for the Elderly – Screening Version (HHIE-S)

- 10-question index to measure social and emotional handicap from hearing loss
- Time = 3 minutes
- Cost-effective
- Sensitivity = 63-80%; Specificity = 67-77%
- Widely-used

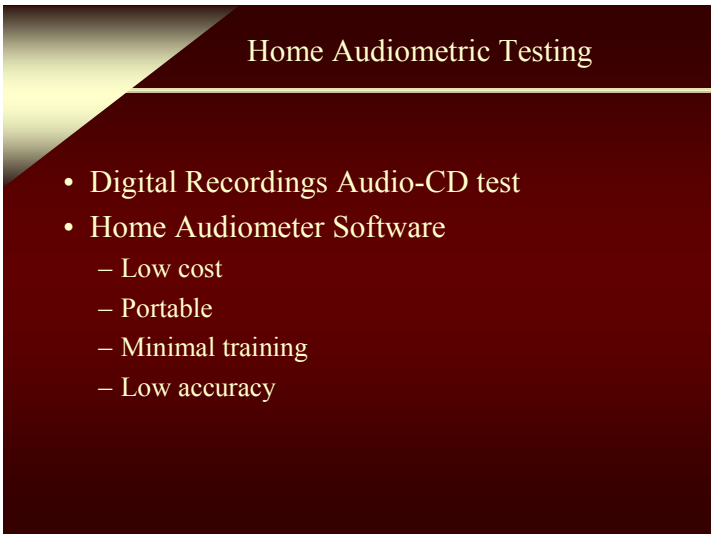
Rather than trying to assess hearing, per se, this screening test indicates whether one's communication, interactions, confidence, and emotional health are affected by a hearing loss. The screener is short and cost-effective. But, as with any sort of self-report looking at sensory function, the sensitivity and specificity are going to be lower. We will likely identify people with moderate to severe hearing loss, but people who are in the early stages of hearing loss may not recognize or admit these hearing problems, so we would miss a lot of people in that mild category. However, the HHIE-S is widely-used in population-based studies, and it is fairly well-accepted in the literature.

Global Question

- "Do you have a hearing loss now?"
- Sensitivity = 78-93%; Specificity = 56-67%
- Brief, cost-effective
- Affirmed as an alternative to the HHIE-S

Another option, which is even broader, is just a single, global question: "Do you have a hearing loss now?" Or there's a rephrasing: "Do you feel you having a hearing loss now?" For this question, the sensitivity and specificity, interestingly enough, do not significantly differ from the ten-question HHIE-S. Some researchers claim that this single, global question could be used as an alternative to HHIE-S, but I haven't seen any studies that rely only on the global question as a measure of hearing. Several studies have incorporated both the global question and the HHIE-S.

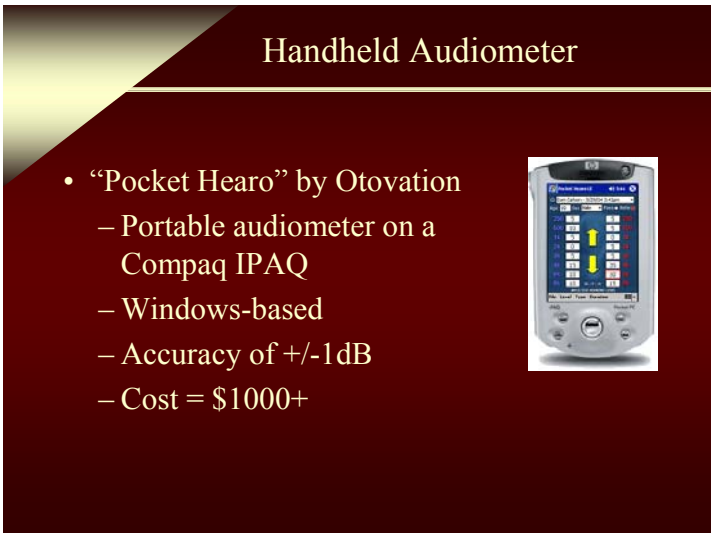
For our study, however, we had hoped to go beyond relying on self-reports. While self-reports have been shown to be valuable in pervious research, they certainly do not equal an objective measure. We are particularly concerned with how the respondent's own perception of their hearing (and other senses) does or does not correlate with his or her actual sensory function. Unfortunately, right now, there doesn't seem to be an appropriate, feasible, and reliable method for measuring hearing in a population-based study such as NSHAP. However, there are a number of different avenues that we found that may be possible in the near feature -- new things that are just on the horizon.



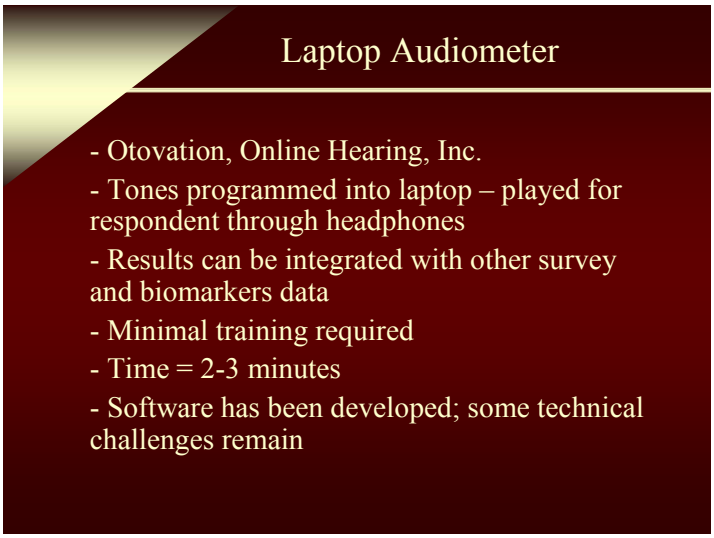
I actually found one of these potential methods by searching on the Internet for home methods for hearing screening. Interestingly enough, because biomarkers are collected in the home, many of the methods for collecting biological and physical measures in population-based research can be found in or derived from home screening methods.

There are two products on the market now for home hearing screening: the Digital Recordings Audio-CD Test and home audiometer software that you can download off the Internet, and run on your computer with headphones attached. With either of these methods, tones are played through headphones (on a stereo or computer) and the respondent marks the tones in a grid, which then enables him or her to derive a hearing score. These products are low-cost, portable, and require minimal training, but the accuracy on these products right now is really low. The audiologists with whom I spoke really cautioned us against using something like this, because there's just no way to know how effectively these methods screen hearing, and how accurately the system – the stereo or computer and headphones – is calibrated.

Another product on the horizon – which may represent the wave of the future – is the Handheld Audiometer. This product is made by Otovation, a company that designs a lot of audiometric products. It is essentially an audiometer programmed into an IPAQ, which is a handheld computer like a Palm Pilot. It has an upgraded sound card, and you simply attach audiometric headphones for listening to the tones. The accuracy is good and it can be calibrated to match other IPAQs – so that tests should be reliable across interviewers. Best of all, the program is Windows-based, so it keeps the data on this small IPAQ, which can later interface with another Windows-based program for managing the data collection.



Unfortunately, the Handheld Audiometer is very expensive right now, at over $1000 per audiometer. We actually talked with this company to see if there would be a way that they could give us a discount, or help us out, or let us get a few of the Handheld Audiometers to do a sub-study, because this really hasn't been used in population-based research yet.



Although that didn't work out, we were, in the process, connected with another company in Canada that has developed similar software that is Windows-based and runs an audiometer on a desktop or laptop computer. It's very similar to what Otovation does on the IPAQ, and it has very good accuracy. Because it is Windows-based, we wondered whether it could be programmed onto the laptops that our field interviewers are going to be carrying with them to conduct the interview. Could we program this on the computer and then run the software to test hearing during the interview? The respondent could put on headphones and hit a key when they hear a tone. Then, the results would be merged with the rest of the survey data. It would be very short, and it seemed like the ultimate solution for us in this kind of setting.

Unfortunately, we got all the way down the line to having our computer programmers talking with their programmers, and we just didn't have sufficient time to get it all interfaced and set up. So, this was our final roadblock. But, we're really hoping with the connections that we've made, and what we see coming along, that this will be something that we might use for our next wave. It presents the opportunity to integrate quick, accurate, and validated hearing testing into a survey or other population-based research. So, that's the future.

In the meantime, NSHAP will administer the HHIE-S and a global question in the first wave of the study. We have also included a couple of questions about hearing aid use. But, we look forward to following up on some of the work that we've started with this and hopefully having a more objective hearing test in a future wave.

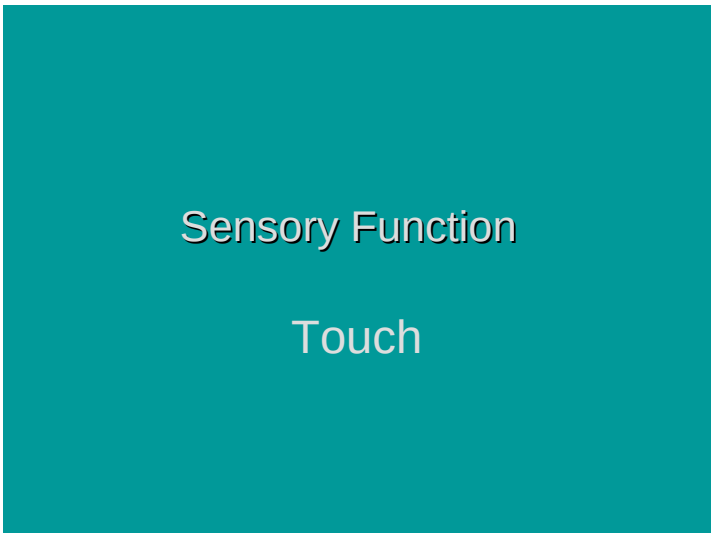
Unidentified Speaker: Very quickly – although this may not be a question for Erin – are you getting veteran status, and whether the respondent participated in a war zone?

Lindau: No, we get their military service, and how long they were in the military, I believe it is, but not specifics about if they were in a war zone with bombs exploding.

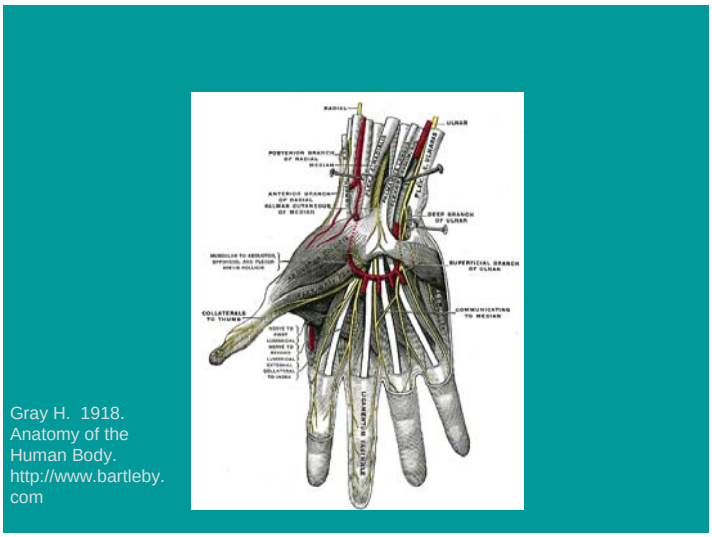
Unidentified Speaker: Okay, because some hearing loss is certainly due to the very high exposures with weapons noise and so on, and everyone trains on it, even if they don't end up being in a war zone. But it would be very useful for a number of reasons.

Sensory Function

Speaker: Sharon Williams



This is basically an overview of what we went through to determine how we were going to test touch. We wanted to test touch for a couple of reasons. Partially because we wanted to round out our sensory function: we're testing everything else, we have to test touch as well, in order to get a good overview of the aging of the sensory function processes that go on. But also because sensory function is important. It's been studied, but not thoroughly in terms of how people age and their sense of touch. It's also never really been looked at in terms of sexual function. And so it was another important area of interest for us. I'm just going to talk a little bit about just some ideas behind a little bit of background, and then I'll go through a little bit more of the methods that we went through to figure out what we were going to use to test touch.



We chose the hand for a couple of reasons. There's been more research done on, say, feet, and sensory peripheral nerves than in the hands, but if you consider the hands as an important communication tool, and that they are important in their exposure to the environment, and they're also something that are going to build up damage over time more than, say, the feet. It's going to be something that's more important in terms of daily function as well, in more, broader activities. So we really wanted to focus on the hands specifically.

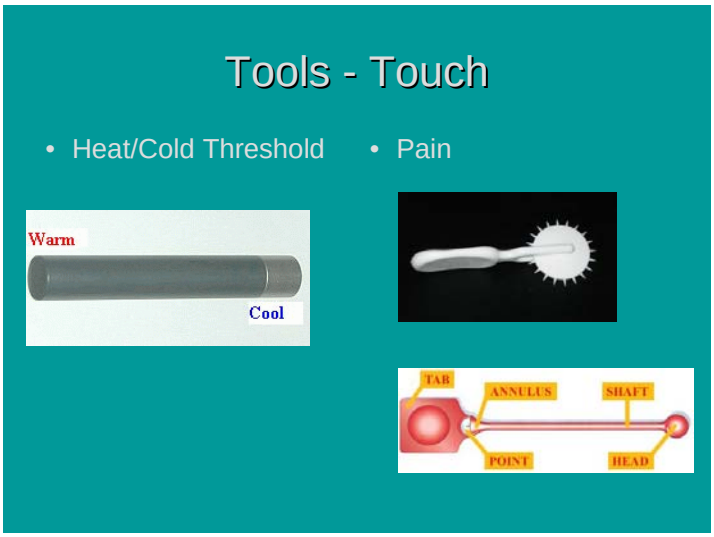
When you consider looking at the hand, touch is only a small part of hand function, all of which are important in terms of both looking at normal daily activities and general overall health, as well as looking at a reflection of pathology of either the peripheral nervous system or the central nervous system. So we were faced with a lot of really big questions about which things we should focus on.

There are three basic areas of hand function to look at. There's strength, both pinch strength and grip strength. There's dexterity, or prehension, the precision movements, and there's also sensitivity to touch. And then, within those broad categories, there're all sorts of different areas to look at specifically.

Our major limitations in this, and I think Erin [York] mentioned these sufficiently, were time and money. I mean, we have three thousand individuals, we will be using approximately two hundred field interviewers, and in order to purchase things for two hundred different field interviewers, something that's going to be expensive is going to blow our entire budget, and it's just not possible. Time-wise, also, we had a limited time for the biomarker section of this particular study, so we were really limited. I think my time limit was something like five seconds. We went a little bit over on that, though not by much.

There're many different methods, again, testing many different aspects of hand function, and tactile sensitivity. Going from the very, very expensive, like the grip dynamometer, which measures grip strength, which runs somewhere around four hundred dollars a piece. It's really the only way, and that's for a very simplified, not-digital, sometimes-hard-to-read, must-be-calibrated instrument. Not practical for our uses. It's really the only way to measure grip strength. You can do things like pinch strength tests, which are a little bit simpler. You can pinch a little piece of paper to see how much you can tug against an individual to get the paper away. There are all sorts of issues in terms of training, and in terms of use of that data as well.

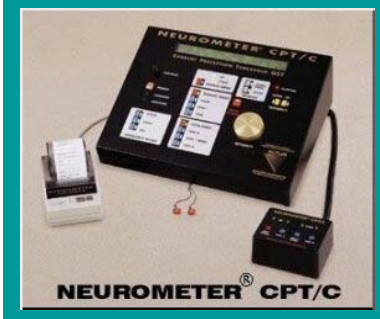
Looking at function and dexterity, there are other tests, such as the Purdue pegboard test, which looks at how long it takes an individual to pick up small pegs from one hole and put them in another hole. Those are a little bit more expensive as well, and they also take several minutes to do, as do the activities of daily living tests, and the Moberg Pick-Up test. The Moberg Pick-Up test is a very interesting one because it combines not only dexterity, but also tactile identification, by picking up small things that are normally found in daily life. Say, a push-pen, or something you would normally find around the house, such as a coin. It is not only the amount of time that you can pick up these items and put them into a container, but also measuring with eyes closed the identification of a specific object. This takes about ten minutes as well. It is relatively cheap, but it does take about ten minutes, which is more time than we were allotted for touch. The activities of daily living were interesting as well, things like buttoning up a sweater, folding laundry, there's a whole list of things that can take from ten minutes to 45 minutes to do this type of test, so it was also not something that we were going to be able to use for our purposes. Therefore, we really narrowed it down to just tactile sensitivity, but there are different measures within tactile sensitivity.



There are basic measures, like pressure thresholds, heat and cold thresholds, and pain thresholds, and they're really looking at separate things. The mechanisms for feeling pain are slightly different than the mechanisms for sensing touch and pressure. So we wanted to also keep this very basic as well, and there are several different methods, running from very cheap to very, very expensive to measure things like heat and cold thresholds, and pain from a simple pin. We weren't really sure we wanted to inflict pain any more than we already were with the fingerstick for the bloodspots, so we decided to go for more of the very quick, very simple methods of touch, measuring touch and pressure, which again, there are many. Another thing we had to take into consideration was using the information that we get for comparative purposes, and, unfortunately, everybody who uses this study on touch uses a different method to actually study touch. So there's nothing that's really used in a lot of population studies.

There are a lot of different things that are used in terms of pathology or recovery after surgery, or something that's going to destroy your sense of touch, but some of the common ones are the use of monofilaments, which are thin wires of different diameters, to actually measure pressure and thresholds, and they bend at a specific point, and to see whether an individual can feel those at different diameters. There are also the Van Boven domes, which are relatively new. They have different numbers of ridges at different distances, and the individual is asked to feel the top of the dome to figure out how many ridges are on top. But there are also problems with these in terms of validity and training, and useful information that's possible, and also time. I mean, all of these take more than thirty seconds, which is what we're looking at.

- vibration



There's also vibration testing, which, if you can just tell by the size of the machine, is not something we're going to buy for two hundred field interviewers to use. So we were down really to an old standby. It's not a new method, it's not revolutionary in any sense of the world. It's the old standby.

The two-point discrimination that's done anywhere from science classes in elementary schools and in occupational therapy labs. Basically, this is either a moving or stationary two points, and you tell at what threshold an individual can actually tell the difference between one and two points. It's not a perfect method. It's the one that we chose because we could do it quickest, we could do it cheapest, and get enough information that we could have useful information from our sample and also for comparative purposes as well.

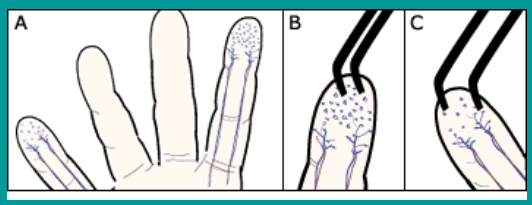
Tools – 2 point discrimination

- stationary

- moving

So we actually chose the stationary two-point test, because for training purposes, it's a lot easier to teach someone how to put two points on an individual's finger than to have them measure a distance, touch it to an individual's finger, measure another distance, touch it to an individual's finger, and continue on with that. It's also a lot quicker. One of the problems with either of these two basic methods is that the amount of pressure you actually put on the finger can determine the kind of results that you get, so you have to be very careful in terms of training, unless you can afford to buy newer innovations in two-point discrimination which have pressure thresholds attached to them, but they're also much more expensive. So we chose the two-point discrimination method for a couple of

2 PD methodology



from: neurotherapy DX –
http://www.ntdx.com/2pd.html

reasons: time, but also because of money. If you look at the actual tools that are available, they're relatively expensive, unless you use the moving two-point discrimination, which we didn't want to use. Also, if you do use the moving two-point discrimination, you have to be very careful in terms of the tool that you use, because the bluntness or pointiness – I don't think that's a scientific term – can determine whether you're feeling pain vs. pressure, and, again, those are two different measures. So we chose a discrimination that was stationary, and we were looking for tools. They're very expensive. The standard, what's called the discriminator (used in a lot of studies), has two discs that range from 16 to 1 millimeter in 1 millimeter increments, and they're about $150. You can get one disc that goes in 2 millimeter increments, that's about 80 dollars. We actually had a shortened version made for us, because I'm not afraid to exploit my family, my retired metallurgist father. Because timing was tight, and also because of the amount of data that you actually get out of these measurements, we wanted to shorten the number, or reduce the number of actual points that we tested. Normal individuals, and there's some decline with aging, can tell the difference between 1 and 2 points at 4 millimeters. That's what's considered normal in clinical settings and population-based studies that have also been done. Most individuals can feel it at 4 millimeters, so we didn't see any benefit in doing the two-millimeter test as well. But also, instead of doing all 8 or 16 of the different tests, we opted for something that would give us still something more than a yes or no answer, but something that would give us a few data points to actually be useful. So we chose three different points. We chose 4 mm, 8, and 12 mm. Those are going to give us some very basic overviews about the sense of touch in these individuals. You know, it's probably not the best method to do, obviously, but with limitations of time and money, this actually works out best for our particular study, and it will be interesting to see how much information we get back from the pre-test, how much variation there actually is. There are a few studies on aging and two-point discrimination, but they're not very broad.

Unidentified Speaker: First, I have an idea about how we might share some of the equipment and things across different studies. For example, the Health and Retirement Study is doing grip strength, and is so the Study of Health, Aging, and Retirement in Europe, so there are quite a lot of these devices around. There's a big upfront cost, we hope that we're going to be re-using them in the future, but there are going to be a lot of empty spots and it would seem efficient to try and use them across different things. And I think that that could go for a number of the other examples that we're getting. That's comment number one. Number two is more of a question, and has to do with this: a lot of these measurements take quite a long time to do, and it seems like it depends on the purpose. Take the last one that we just talked about, about the various points of touch. Suppose you wanted to know about what the relationship between the distance apart and the points of touch over the whole range is? Well, that would be expensive to get on a given individual. If you wanted to only know that on the level of groups, say, age groups, you could randomize the number of measurements, but have varying measurements for each individual. As you get in to studies that would involve a number of different measurements, that is, you wanted to look at hearing and other things, then it seems like you'd need to think about the design, but it does seem to me that that's some guidance from good statisticians. I think that might be useful to figure out how you could measure more domains in a fixed amount of time.

Unidentified Speaker: In response to the comment about sharing, I think one of the things we've been thinking about, with this clearinghouse of information is also, 'Could there be some type of clearinghouse where we can share equipment?' I said yesterday, like the eBay of biomarker equipment. We're hoping that we, maybe even between our studies, we can do that sharing and set an example for others. We, certainly, are working towards maintaining our equipment so it can be used for future waves and also shared.

Unidentified Speaker: Just a comment, and that is that monofilament, tuning forks or vibration perception thresholds have been used for years, simple measures for diabetic coagulopothy, and I'm surprised that in large studies that there hasn't been interest.

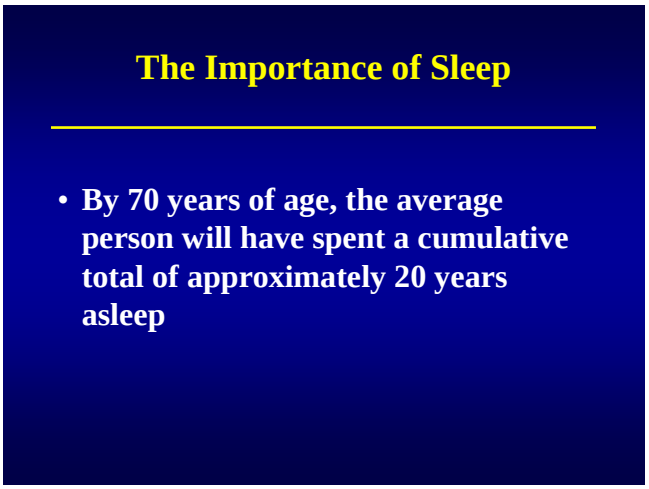
Williams: We did look into tuning forks, not just for vibrations, but for hearing, and it is my understanding that they are prohibitively expensive, at least for our project.

Unidentified Speaker: And also, for training purposes, for field interviewers, it's very difficult to do.

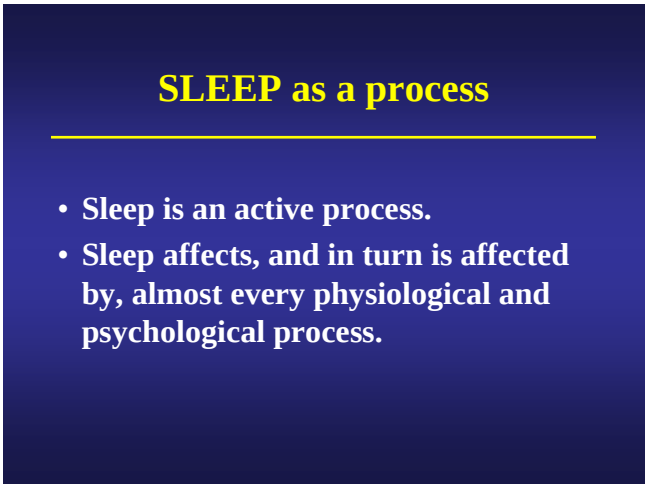
Williams: The monofilaments are inexpensive. You can get disposable ones that are pretty inexpensive, but there's a training issue with that, and also in terms of the time it would take to do one of those versus the normal number that are in fact used. .

Sleep, Chronobiology, Neuroendocrinology Laboratory

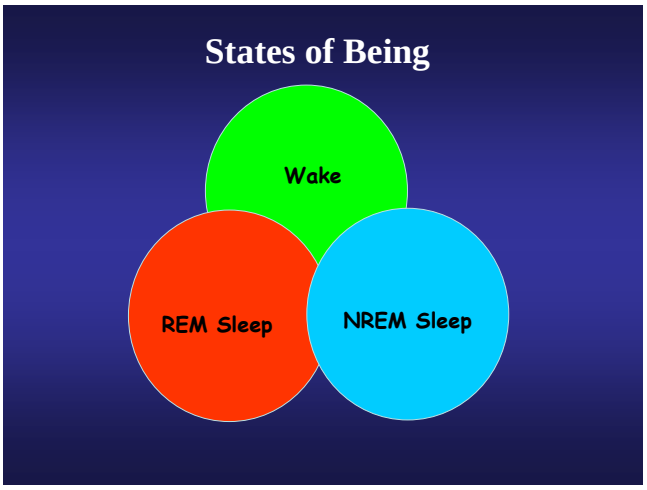
Speaker: Federica Latta



You probably know that by seventy years of age the average person will have spent approximately 20 years asleep, and you probably all know that sleep is not a passive state.

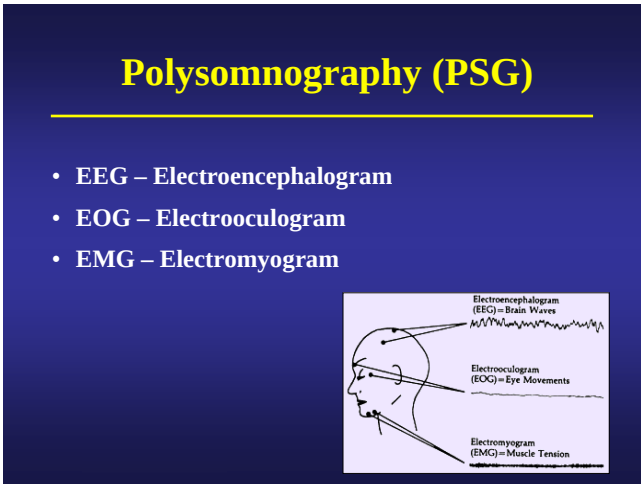


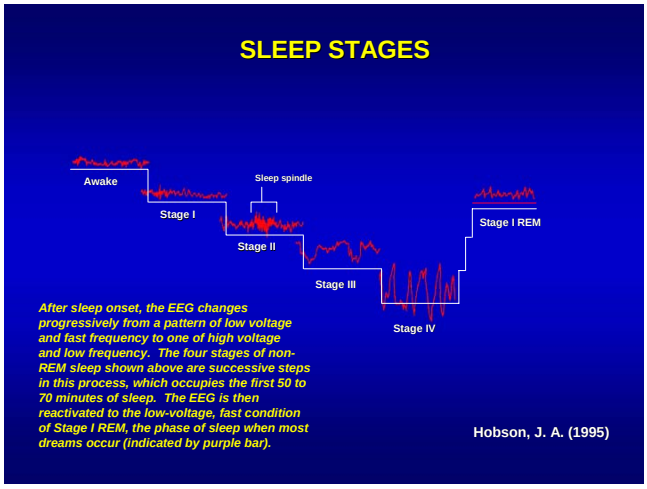
The brain is active while you are asleep, and sleep is actually an active process. Sleep affects, and in turn is affected, by almost every physiological and psychological process: sleep not only affects your psychological well-being, and your cognitive performance, but also your physical health.



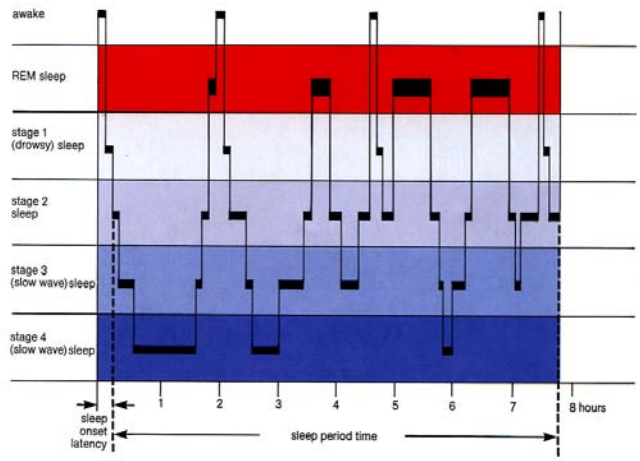
It has been suggested that there are three states of being: wake, REM sleep (REM stands for rapid-eye-movement), and non-REM sleep.

The gold standard to record and to measure sleep is polysomnography: you have the electroencephalogram to record brain activity, brain waves; you also have electrooculogram to record eye movement - this is important for the REM sleep, the rapid eye movement sleep; and electromyogram to record muscle tone (usually in the chin).





Here you can see the signals that you get from electroencephalography: after sleep onset, the EEG's patterns change progressively from a pattern of high frequency and very small amplitude that you see in wakefulness, to a pattern where you have high-amplitude and slow frequency waves. As you can see, there are different stages. You go from wakefulness to stage-1 (sometimes also called drowsy sleep), to stage 2, stage 3, and stage 4. These are criteria defined by Rechtschaffen and Kales. Here you see the rapid eye movement sleep (REM sleep), in which the EEG signal looks very similar to stage 1 or wakefulness.



This is the information that you can get from polysomnography. This is a sleep period of approximately eight hours: the person usually starts being awake, and then from wakefulness, the person enters stage 1, and then stage 2, stage 3, and then stage 4. These two stages -- stage 3 and 4 -- are the deeper stages of sleep, because it would be harder to wake up a person that is in deep slow-wave sleep, with the big waves that I showed you earlier. Stage 3 and 4 are also called deep sleep. And then as you see, from stage 4 you go back to 3, and then 2, and so on. This first cycle takes approximately 90 minutes and repeats itself through the night. It's important to notice also that the first half of the night is not the same as the second half. In the first half of the night, you have a lot of deep sleep, that is stage 3 and stage 4, but less so in the second half of the night. If you look at REM in red up here, you see that the first REM episodes are much shorter than the ones in the second half of the night. The early morning is when most rapid-eye movement sleep occurs.

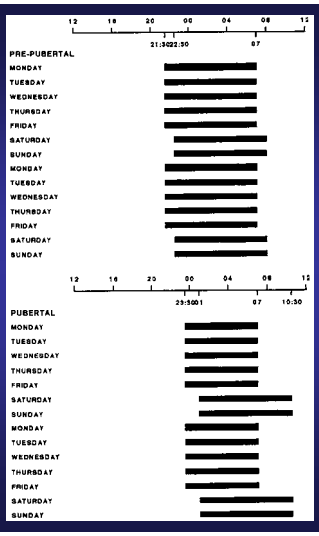
Recording the rest-activity cycle in the human

- Self-report
- Wrist actigraphy

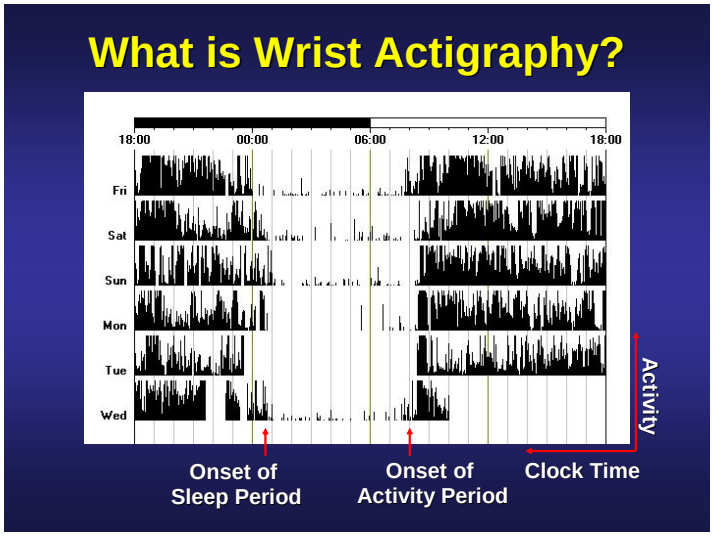
How do you record the rest-activity cycle in humans? Of course, you can have self-reported measures: you just ask the person to record bedtime and wake-up time. Obviously, that has all the problems of self-reported measures, and I have here an interesting example.

The time is in military time. At the top, data from a pre-pubertal individual are plotted. Usually bedtime schedules are pretty regular during the weekdays, and in the weekend, people go to bed a little bit later and sleep in a little bit later. And here, you see the shift of the biological clock that is typical of a pubertal individual. This shift in the biological clock of adolescents has received a lot of attention recently and some schools have decided to start later in the morning because of sleep-deprivation being so common in children and teenagers, and students falling asleep during the earlier classes of the morning.

Example of data obtained using self-report



The other way of getting sleep/wake data is through wrist actigraphy.



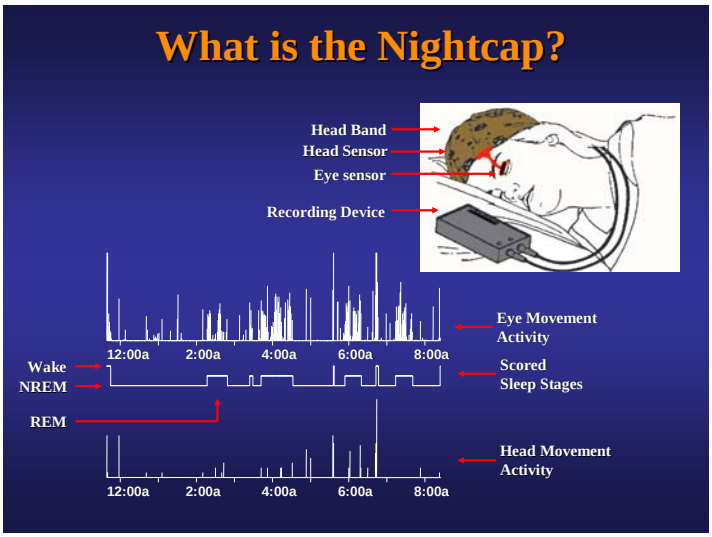
These are the actual data that you will get from an individual recording. Each line is a day, starting with Friday, then Saturday, Sunday, and so on. You get these data through a very light device that looks like a wrist-watch. It's very light to wear, and it records wrist movements. If you're playing tennis, you're going to see a lot of wrist activity while you are playing. If the person is asleep, there is much less activity. And here you can see that this particular individual went to bed at around midnight, Friday night, and woke up, I would say, at 7:45. As you can see here, there is still some activity during the nighttime. These are data that come from a young, healthy individual, and the amount of the activity is quite low during the nighttime. If you were to measure wrist activity in an older individual, you would see more activity during the night because older individuals have more fragmented sleep, more awakening, and more posture shifts.

Measuring sleep-wake cycles

- **Polysomnography at home or in the laboratory (gold standard)**
- **Questionnaires**
  - **E.g.: The Karolinska Sleep Log**
- **The Nightcap**

The gold standard for recording sleep is polysomnography. It can be done in the sleep lab or in the home of the person, but as you can imagine the home recording is quite expensive, since you need a sleep technician that goes to the subject's house to hook up all the electrodes, and you need a portable computer that records the night of sleep. And then you need this technician to go in the morning to remove the electrodes, and also to bring the computer to the laboratory so that you can download the data. It's costly and it's probably not ideal, because the computer and the equipment are quite heavy.

Questionnaires and sleep logs are widely used. A sleep log can ask you not only at what time you went to bed and what time you woke up, but also how much time it took you to fall asleep, and how refreshed you felt in the morning when you woke up, how soundly you slept, and how well you slept. For example the Pittsburgh Quality Index questionnaire asks you questions regarding the quality of your sleep in the past month.

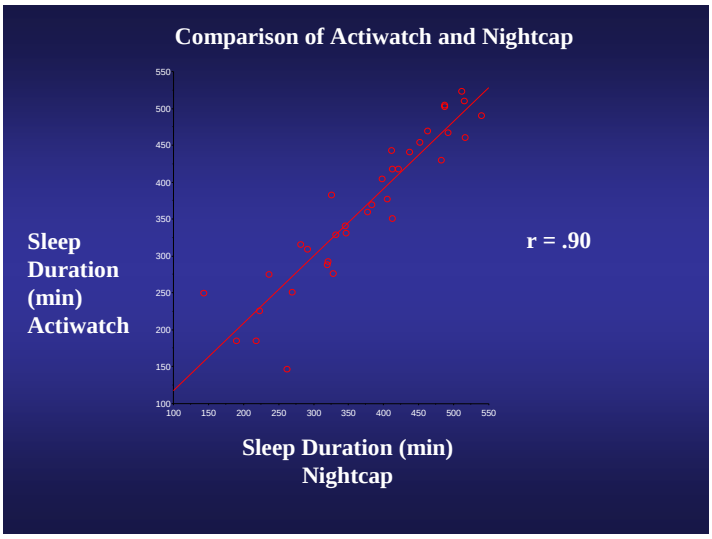


Another interesting device is the nightcap, and I am going to show it -- I have it right here. Basically, you have a headband or a bandana. Then you have a head sensor that you can see here, and an eye sensor that is something like that, where you can attach this sticker, and you put it in the eyelid. This can be done by the subject himself at home just before going to bed: the person just has to position the headband in the right way and peel the sticker and put it on the eyelid, and this is the recording device. I myself have tried this at home, and you can sleep comfortably and you don't even notice that you have this on your head. You get the eye movement activity here on top and you have the head movement activity at the bottom. This functions on the principle that if you are both moving your head and your eyes, you must be awake. And if you are not moving your head and you are not moving your eyes, you must be asleep. But if you are moving your eyes without moving your head and your eyes are closed, you are in rapid eye movement sleep.

Measuring Sleepiness

- **Subjective scales**
  - **Visual Analog Scales (VAS)**
  - **SSS**
  - **Epworth Sleepiness Scale**
- **Cognitive Performance**
  - **PVT**
- **Objective testing**
  - **Multiple sleep latency testing (MSLT)**

This is different from wrist actigraphy because with the nightcap you can distinguish not only between wakefulness and sleep, but also between non-rapid-eye-movement sleep and rapid-eye-movement sleep. Therefore, the nightcap gives you a little bit more information about the sleep architecture. With polysomnography, obviously, you don't just get time spent in bed, but you actually get how much time the person was asleep during the night, and you also get sleep fragmentation, sleep quality, sleep efficiency, and so on .



This shows a comparison of the wrist actigraphy and the nightcap, in measuring sleep duration (in minutes), and you can see that the correlation is quite good.

Another thing that you can do is to measure sleepiness. In order to measure sleepiness, you can use subjective scales, questionnaires, or you can do tests of cognitive performance. The gold standard is the multiple sleep latency test, but this needs to be done in a laboratory, with a polysomnography machine and a sleep technician who is actually looking at a computer screen to detect the sleep onset. Regarding the questionnaires, we have used in our laboratory a visual analog scale for global vigor.

Global Vigor - VAS (Visual Analog Scales)

How alert do you feel?
very little ——————— very much

How much of an effort is it to do anything?
very little ——————— very much

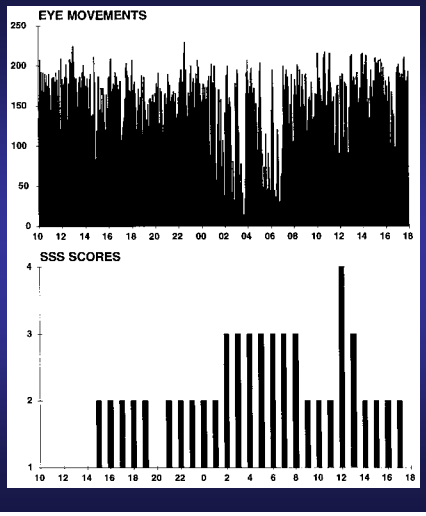
How weary do you feel?
very little ——————— very much

How sleepy do you feel?
very little ——————— very much

You are probably familiar with this: the line is exactly 10 cm long, and the person has a paper questionnaire, and puts a mark with a pen along the line, corresponding to how they feel. For example, now I'm very alert, and I would put the line almost at the extreme.

Another common questionnaire is the Stanford sleepiness scale: the person is given this list, on a sheet of paper usually, and the person has to rate how sleepy he or she feels. It goes from feeling active and wide awake (1) to almost losing the struggle to remain awake (7). We usually administer this questionnaire on an hourly basis, because we want to get the circadian rhythm, so every hour the person says 'I am a 4,' or 'I'm a 5,' and it's quite efficient and quick.

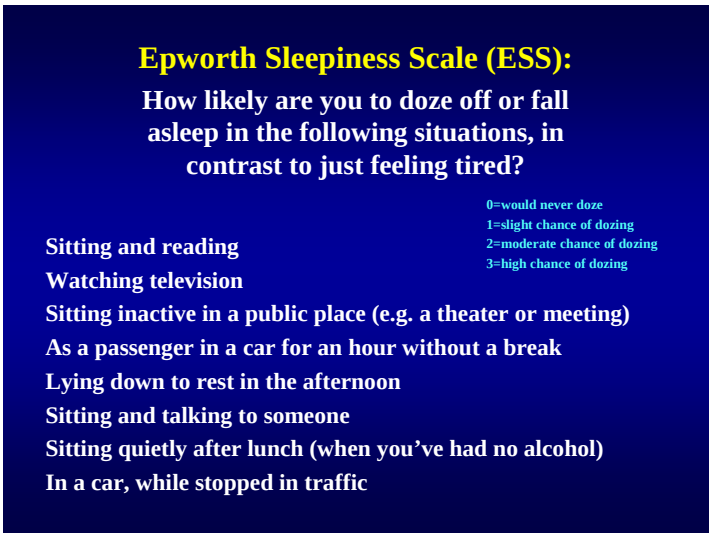
The Nightcap can track sleepiness



Stanford Sleepiness Scale (SSS)

1. **Feeling active and vital; alert; wide awake**
2. **Functioning at a high level, but not at peak; able to concentrate**
3. **Relaxed; awake; not at full alertness; responsive**
4. **A little foggy, not at peak; let down**
5. **Fogginess; beginning to lose interest in remaining awake; slowed down**
6. **Sleepiness; prefer to be lying down; fighting sleep, woozy**
7. **Almost in reverie; sleep onset soon; lost struggle to remain awake**

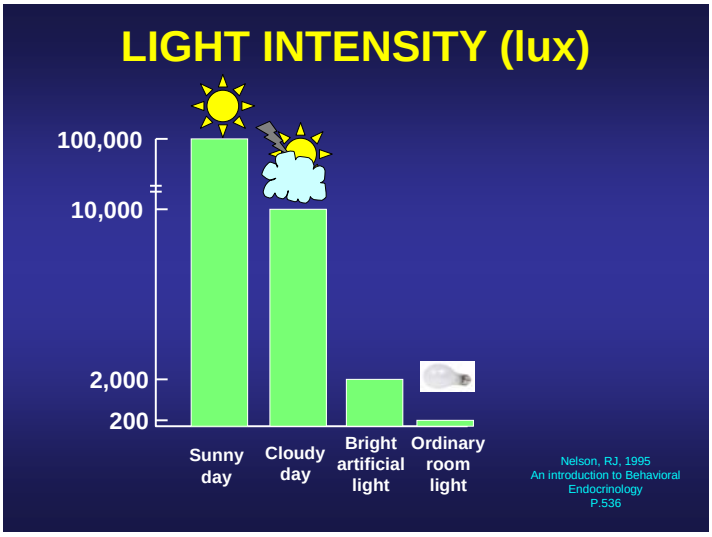
X. **asleep**



Another interesting questionnaire is the ESS, which measures excessive daytime sleepiness. Basically, the question is, 'How likely are you to fall asleep or doze off in the following situations, in contrast to just feeling tired?' you have several situations: sitting and reading; in a car while stopped in traffic; sitting and talking to someone. And the person has to answer, checking a number. The score goes from 0 through 3 for each item or each situation here. 0 is "would never doze," and 3 is "high chance of dozing." I've given this questionnaire to students, sometimes in classes, and I was amazed to see the responses that I got, how high the scores of students are. It's almost scary to imagine that there are people that are so sleep-deprived that are walking around and driving around, and they are with us. But if you have ever been in a classroom, even a seminar, I'm sure you have been exposed to people sleeping in all sorts of positions.

Another very useful test for alertness is the PVT, which is very conveniently available on a Palm Pilot. It's sort of a reaction-time test, and the good thing about it is it has no learning curve, so you can start doing it and you can use the data right away. It's very simple, and it consists of this: a round black dot like this appears on the screen, and as soon as the person sees this, they have to press this button. It can be adjusted for a right-handed person, or a left-handed person. And it stores the data. Usually this test is administered for ten minutes. It can store the data for up to twenty tests, and then it calculates the statistics, the mean reaction time, how many lapses, minor lapses, major lapses, and then at the end you can download everything into a computer.

One thing that I want to point out to you is also how important light intensity is for alertness and sleepiness. If you think of light intensity as measured in luxes, you can see that on a sunny day, usually you have 100,000 luxes. In an ordinary room, there are about 200 luxes. There is a huge difference, and this is especially relevant if you are thinking about studying older adults that, because of mobility or other problems they might have, like less social activities, are less exposed to outdoor light. Elderly adults, especially in nursing homes, spend their days in very poorly lit rooms, and at nighttime, they are never completely in the dark, for example to facilitate nurses. Elderly in nursing homes are never sleeping in the dark, and in the daytime, are never exposed to bright light.



Biomarker Collection in Population-Based Health Research: Human Subjects Issues

Biomarker Collection in Population-Based Health Research: Human Subjects Issues

Speaker: Kathleen Mullan Harris

For this session, we were asked to think about the major challenge for human subjects regarding biological specimens. My answer to that issue would be archiving the biological specimens, and that's an issue we are facing in Add Health, a study that I am now directing. In Add Health, we view collecting biological specimens no different than some of the other very sensitive survey data that we've collected from the beginning. I thought that I would take a couple minutes to describe Add Health, because the design is what determines the level of sensitivity that makes us panic about human subjects issues. Then I'll describe our security system.

Add Health, for people who aren't familiar with it, the long name is National Longitudinal Study of Adolescent Health, is a national longitudinal study of adolescents in grade 7 through 12, in the United States in 1994 and 95. It's a school-based design, where we selected 80 communities, and within that design, 80 high schools, and then 80 feeder schools, or middle schools that feed into those high schools. In the school administration, we went into those schools and we interviewed all the students in those schools on a particular day. This came to about 90,000 students who participated in the in-school administration. Then from the school rosters, we selected a sample of adolescents for in-home interviews, and it is this sample that represents our longitudinal panel. From the school rosters, we selected about 200 students from each school pair, and also selected a number of special oversamples, that came to a sample size of about 21,000 adolescents and their parents. We then conducted a wave I in-home interview with adolescents and their parents. We've completed three waves of in-home interviews. The most recent one occurred in 2001-2002, when our respondents were 18 to 26. In the first two waves of data, we only collected survey and anthropometric data, and in the last wave we collected biological specimens. We collected urine for testing of STDs, saliva for testing of HIV, and on a sub-sample of our genetic sample, which I'll explain in a second, we collected buccal cells for DNA.

Now, why did we worry so much about security from day one? Because our design creates a highly clustered sample. The design was developed to understand how the social context of adolescent life affects adolescent behavior and health status. So we have contextual data from the school; we also interviewed the school administrator. We have contextual data from the family – we have the parent interview, but we also have a nested genetic sample, which contains 3,000 pairs of adolescents who have some genetic resemblance. For our genetic sample, we oversampled identical twins, fraternal twins, full siblings were not oversampled because they occurred naturally in lots of numbers, we oversampled half-sibs, and then we oversampled adolescents who were biologically unrelated but live in the same household. In Add Health there are about 500 adopted kids.

We have neighborhood data and community data, which were merged using extant sources because we collected the respondents' addresses which were geo-coded. And in the in-school administration, our respondents nominated their 5 best male and 5 best female friends, and these friends were, for the most part, also in the survey, (because they were, for the most part, school friends). So we're able to link the friends' responses to the individual who nominated them. They nominated friends off of the school rosters. They also nominated their romantic partners and their sexual partners, and these are referred to as third-party nominations.

The other thing about Add Health is, because it was school-based, a lot of people knew other people who were in Add Health. We figured out that if you add in the parents and siblings and other family members, it was about 360,000 people who knew somebody else who was in Add Health. So as you can see, widespread and local knowledge about who was in Add Health are the human subjects concerns that really worried us. We therefore developed an innovative security system, and there were two principle risks that we were worried about. One, of course, is the breach of confidentiality, and then the second was the risk of deductive disclosure, because of Add Health's highly contextual design. The key to our security system, and we've been talking about this, took years of discussion and angst and worry among project investigators and staff. It takes a long time, I think, to work through a lot of these issues with a particular study.

Anyway, the key is that we separate respondent identifiers from the data immediately upon finishing the interview, and have set up an honest broker system from the beginning. This will lead into the next session. What is an honest broker? He or she is an intermediator between the subject, the data, and the researchers. This intermediator is what we call our security manager. The security manager is located outside of the United States, and our security manager holds all Add Health identifiers. We have no identifiers of our respondents in the U.S. So by identifiers, we have names and addresses, and that also includes Social Security numbers and other family identifiers. We at Add Health have no idea of the names of any of our respondents in the study, or where they live, and can not get that information. We don't release date of birth or the geo-codes.

Unidentified Speaker: Do you keep date of birth and geo-code in your data set?

Mullan Harris: Only momentarily. Only for moments in time when we're collecting the data in the field, and cleaning data files for public use. Once files are clean, these data remain with our honest broker.

Unidentified Speaker: So you destroy the intermediate analysis data set before you get your final analysis?

Mullan Harris: Right, right.

Unidentified Speaker: That's sort of what we do, too.

Mullan Harris: We're accused of being paranoid. One, because of our immediate delinking of identifiers from data while we're still in the field, and second because we require security plans from users when we disseminate Add Health data. It's more with dissemination that it's difficult for public users to understand all the security requirements that we put them through. But most people are really pretty understanding. The security manager keeps the identifiers, and the traces.

Unidentified Speaker: Does your honest broker actually merge it together, if you have data coming from different sources? This is so-and-so's data from the interview, and this is the data from the lab, and they put it together, and then what you get is a dataset that's merged with no I.D's on it anymore?

Mullan Harris: Yeah, that's exactly right. It's an elaborate matrix of I.D.'s. So, for the moment in time our field contractor– who was NORC for the first waves of

data, and was RTI for this last wave – has the identifying information of our respondents, they follow our security procedures. Here is what happens. The field contractor has a survey, and they have the names and addresses of respondents, and then they connect a survey I.D. with the wave of survey data, and also a specimen I.D. for the particular survey respondent. They conduct the interview, and at the end of the interview, the identifiers are stripped off. You basically create two files on the computer. The identifiers and the addresses are in one file, along with the survey I.D. and the specimen I.D., and that gets sent to our security manager. And then the other files contain the survey data and the survey I.D. The biological specimens have their own specimen I.D., and they get sent to the lab. Then, as soon as our security manager says that they've received the file with identifiers and survey and specimen ID's, our field contractor destroys that file of identifiers, so it's no longer in our hands

Unidentified speaker: And it's sort of similar to what we do in our own little way.

Mullan Harris: So that's how that works. The survey data get cleaned, and get sent up to the security manager, the biological specimens are analyzed and assayed, results are sent to the security manager, and they bring everything together, and then they create what's called an alternate I.D., or an A.I.D., for the file that merges everything together. They send that to us, and then we disseminate those data. Therefore, the A.I.D. is what allows people to merge data over waves. That's all I was going to explain about our security system. In wave 3, when we collected biological specimens, we already had the security system in place, and that took care of the confidentiality of our biospecimen data and their results.

The issues that we were most worried about were, in terms of human subjects, third-party nominations, because our respondents nominated other people and then reported to us a lot of information about them, and very sensitive information in terms of sexual behavior and illegal behavior. And we don't have the consent of these third-party nominations. That's really what makes us nervous. With the biological specimens, at least we have the consent, and you can make the pledge to them that nobody will ever be able to connect biospecimen results with your name. This is what we view as very sensitive. Also, the design informs you about what is sensitive. So, for example, in wave 3, we have a couples sample, where 1,500 of our respondents recruited their romantic partner to become part of the Add Health sample and participate in the interview, and then we tested their biological specimens for STDs. We have very low rates for certain STDs, especially for HIV, so releasing these data on couples where you've got test results on partners increases the motivation for a person to try to break into the dataset and figure out who's who and figure out what those results are. Those are the kinds of issues that you need to think about, and the design determines the level of risk of disclosure and potential data intruders.

I can just end with the biological specimen protocols we used. In wave 3, the consent forms were pretty general. We archive the urine, and then we accept proposals for future analyses of archived urine. So our consent form said that

we were going to keep the urine. They were very general, and I'm not sure that will be what we'll be able to do next time. Especially because we plan to collect DNA on everyone in wave 4, and so those issues that we'll be getting into with archived DNA is a different ballgame for human subjects. Maybe I'll just stop there.

Unidentified Speaker: Who gave consent? Did you get consent from all the parents?

Mullan Harris: In the first two waves, we had parental consent. In wave 3, all of our respondents were over the age of 18.

Unidentified Speaker: I have a question about these third-party participants. Are they interviewed in any way, or are data collected for them? And if not, do you really need their identifying information?

Mullan Harris: Some of the third party nominations would be interviewed. For example, the friends that are nominated. Most of them were in the in-school interview, and then they became part of the in home sample at a rate of about 2 out of 9. Although they gave consent to participate in the interview, they didn't give consent to have their friend report on their behavior. I think that was the issue. I think the IRBs are cracking down on the third party nominations.

Unidentified Speaker: You were talking about archiving of biological urine, which is something that we're looking into what our IRB requirements are for storing specimens for future analysis, and you mentioned it was general. One thing that came back from our IRB is that we need to inform respondents in the future if we're going to conduct future tests on that specimen. Is that something that you've had to deal with on Add Health?

Mullan Harris: No, this is one of the things that we thought long about. We did not have to do that in wave 3. We told our respondents that their urine was going to be tested for STDs, and told them which STDs: chlamydia and gonorrhea and HIV, and we actually reported the results for the curable STDs – Chlamydia and gonorrhea – by using an anonymous call-in with a pin number; and the same for HIV. But we also tested them for trichomoniasis, for which there is no cure, and we did not report results on that. Then we said in the consent form that we would conduct, 'other possible tests.' It was extremely general. We had incredible compliance, 92% of our respondents consented to provide urine, and 95% of our respondents consented to provide saliva for the HIV test. For future testing though, as you see, we can't go back to our respondents for consent, that's something we just can't do because we break the link with the identifiers immediately while still in the field, and this procedure was very hard for our field contractors to live with. Still, it's our security system and that's just the way it goes. If the field contractor makes a mistake in the field, they can't go back to the respondent to fix it, and we lose that case. In the future, if there were some medical relevance that we could ascertain with our archived samples, if, for example, someone came across a gene that determined some kind of life-threatening disease and you could test for that in our DNA, and the IRB would require a consent, we would have to pass on that. I think that would just be something we'd have to live with.

Unidentified Speaker: How did you deal with HIV reporting to states?

Mullan Harris: We didn't have to report because we were covered by our certificate of confidentiality. If we would have had to, we wouldn't have collected it.

Unidentified Speaker: It seems to me that it would be really useful to have a network of IRB people involved in making decisions about large population-based research studies involving biomarkers, because again, just like we do with the protocols for the biomarkers: we reinvent the wheel, which is such a waste of resources. I wonder if we can think, certainly Add Health, HRS folks, NSHAP and other large national studies, how it is that our IRB folks might interact?

Unidentified Speaker: Yeah, I've been worrying about this for a long time, and the issue of multiple IRBs with community members is a concern. I think really that the wave of the future is one national IRB for some of these large studies, and there is a national IRB for cancer. I mean, if an IRB has membership for two or three years with people trained, it can review these studies on a consistent basis, and it'd save everybody a lot of trouble. I don't mind going to an IRB and going through a full review, a careful review once. But going through it thirty times is not only a waste of our time, it's a waste of their time. I think consistent guidelines would come out of a national IRB, too.

Biomarker Collection in Population-Based Health Research

Speaker: Parminder Raina

Thanks very much for inviting me to this meeting. I will provide a quick history and background about the Canadian Longitudinal Study on Aging. We are still in the development phase of this study. Many of the issues that have been discussed here are similar to the issues that are being considered by the CLSA team. Some of these issues are related to ethical and legal issues and others are more methodological in nature. We're doing some feasibility studies to resolve some of the methodological issues including ethical and legal issues around the launch of a population-based study in Canada. But before I talk about those specific issues, let me quickly give you an overview of what we are planning to do. It's a work in progress, so bear with me. Let me also acknowledge my Co-PIs, Tina Wolfson, who was here yesterday, and had to leave. Then, Susan Kirkland, from Dalhousie University, she's the other Co-PI on this study. We also have 200 co-investigators across the country.

In just a thumbnail sketch, what are the overall aims of the CLSA? We're interested in looking at aging as more of a dynamic process. We are interested in looking at interrelationships among intrinsic and extrinsic factors, and we are looking at it from a midlife to old age perspective, from an adult development perspective. Even though we will be capturing lots of health conditions, disease, and lots of other disability issues in this study, our overall main focus is looking at healthy and successful aging.

One of the most important functions of this study is to serve as a platform for future research, and also to build capacity in the area of aging and health in Canada. Just to point out that this is actually not an investigator-driven initiative, it is an initiative of the Canadian Institute of Health Research, which is the equivalent of NIH in Canada. So our job is not to do any fund-raising: CIHR is doing it independently of the development of the science. The conceptual framework that we've pursued, as I've said, is health and successful aging, adult development. We are trying to capture some of the life course by asking some of the questions from the individuals about their life history.

Adaptation. We are very interested in knowing why some people age successfully and others don't. And complexity. This is one of our underlying principles. Looking at integrated questions that look at bio, psycho, and social aspects of aging. These are some highlights of different areas where we are collecting information. For physical functioning we are looking at disability, acts of daily living, frailty, co-morbidities, injuries, a whole host of chronic diseases. In the psychological functioning area, we are interested in looking at cognitive functioning, values, and meaning. Everyday competence, adapting functioning, coping, personality, emotion, psychopathology, and psychological distress.

Under the social functioning, we are proposing to look at social networks and social support. Work-to-retirement transitions, structural inequalities, matters of place and mobility, and so on and so forth. Then we have what we call an inter-theme content, where we're looking at some of the biomarker issues, which span different questions that might come up within this study, and genetics of aging. Within that, we are proposing to look at genes of longevity, DNA repair, anti-oxidant defense, apoptosis, programmed cell death, immunosensence and teleomere loss.Then, we have lifestyle, health services, quality of life, pain, and spirituality.

Currently, we are in the process of refining the content of the CLSA to provide focus and coherence. We went through international peer review for this particular study, and generally people liked the idea of what we're trying to do, but obviously we need to focus a bit. The design of the study is a longitudinal design: we are looking at men and women over the age of 40, and the sample size is 50,000 individuals. We plan to follow individuals for 20 years. The repeated measurements for people between the ages of 40 and 79 will be every three years and for people over the age of 80, it will be every year. Also, as a part of the study, up front, we are building some embedded studies. We are also doing substantial linkage to existing databases. In Canada, we have opportunity-linked data to health care realization databases, some of the disease registries, and some of the social databases. It raises some interesting challenges for us and eventual public access of data to the larger research community. But that raises a whole host of issues around security and confidentiality, and who gets the data and who doesn't get the data. As we were developing this, one of the things that we had to go through was a reality check around the feasibility and the cost of the proposed study, and so we actually ended up dividing our study into two cohorts. First, because this study has to be relevant to the policy makers, who are going to be funding a lot of this, and second, it has to advance the science of aging. So we have actually created two cohorts. One which we call comprehensive cohort, where we are doing extensive testing, which I will talk about a little bit more. And then what we call tracking CLSA, which is computer assisted telephone interviews. That's a nationally representative sample that can give provincial level estimates for different questions that might be of relevance to provinces and to the federal government. Comprehensive is only going to be done in six sites. It's a national scope, but not nationally representative. It's going to be done at six academic sites; they will be recruiting subjects within a 100-km radius. For the comprehensive data, we are proposing to collect fasting blood samples, urine samples, and skin cells. Some of our biologists want to do skin punches on individuals, which actually require two to three stitches after the skin punch is done.

So we had a lot of debate about whether we could do it or not, and that's something that we're actually going to do a public consultation across the country to see if people are going to be able to provide those skin cells. It has a lot of scientific value, but can we do it? That's a different matter.

Unidentified Speaker: I've done it. Just little core biopsies.

Unidentified Speaker: You've done it in the context of a large research project?

Unidentified Speaker: In one community, and people might or might not be at risk for certain diseases. It's not a national survey.

Unidentified Speaker: But you did this procedure in the clinical setting?

Unidentified Speaker: In the homes.

Unidentified Speaker: Just a punch biopsy, not suture?

Unidentified Speaker: Not suture.

Unidentified Speaker: All right.

There are certain issues about standardization, about the storing of the data: where it gets stored, where do we analyze data. Because of the size of the study, you also need labs that can handle the volume of the sample that comes through, and we are fortunate that at McMaster University we have a clinical

laboratory that handles over 100,000 samples a year from some international studies, so we are proposing to use the MacMaster lab to store, ship, and do everything related to our samples. Then there are some specific tests around the genetics that we will be recruiting our co-investigator labs to do. I'll just quickly take you through the sample selection we are doing: blood, urine, and skin.

From the blood, we have blood cells, serum, and plasma. From blood cells, we are doing different types of genetic-type preparations that might be used for future research. A lot of it is going to be stored, very little is going to be analyzed right away, and that raises a whole bunch of issues around ethics, consent etc. Then there are epithelial cells and fiberplast cells that people are interested in looking at. I will skip the tracking cohort.

As I mentioned already about the data linkage, we are proposing linkages to administrative databases, disease registries, and with some of them, there is a protocol that already exists, and with some of them, we have to go out and figure it out. We are also doing some macro-level data linkages to environmental data, so we are looking at not only individual-level information but also the systems-level information.

What are our challenges? There are the implementation challenges. For example, one of the proposals is to look at fasting blood samples. I know lots of studies have done this, but not for a study of this magnitude, in relation to, for example, home interviews with individuals. At that time, we were thinking that we would ask them to set up an appointment to come to a clinical setting, where they will go through a clinical evaluation and give the blood. But, for the fasting blood, that means we have to do all our clinical evaluation in the morning because of the fasting and it would be difficult to keep people around for a whole day. We are looking at those challenges, how we are going to deal with it. Should we split the blood collection from the clinical evaluation of the individuals? And what are the best ways to collect blood samples? Do we do home visits, do we recruit private labs, do we have CLSS-specific labs, or do we use hospital labs? Also, the ability of these different structures to prepare samples and ship them. CIHR has also established an ethical, legal, and societal issue (ELSI) committee, which is comprised of lawyers, ethicists, geneticists, epidemiologists, social scientists, and privacy commissioners. Now, the debate is what is legal and what is acceptable in Canada. We are having a lot of debate about informed consent, especially about ability to consent. Yesterday, we had a bit of a discussion about cognitive capabilities of individuals, and whether we should include cognitive-impaired individuals. Some of the lawyers and some of the ethicists talked about if people who are depressed are able to give consent? So it sort of becomes of a larger issue.

Issues of proxy consent are also being considered by our team and ELSI. What happens when the individual becomes cognitively impaired during the study? Does that proxy consent, which was given ten years in advance, hold true or not? Full consent vs. staged consent. Based on the discussion within our ELSI committee, it will be a long consent form. Do we complete the whole consent form at the beginning of the study? Or do we take them through stages, as the study progresses? That has its own ethical and legal issues: Genetic and biochemical and the future analysis of this data.

One of the challenges that has come out of our study, that we haven't found a big example of anywhere else, is the whole commercialization issue. Do we tell our participants right up front? There is no general guidance on how to deal with the issue of commercialization. There's a very strong feeling that we should, but what impact it has on the participation in the CLSA, revenue generated via these commercializations. And also, the issue of blanket consent vs. issue-related consent. Do we ask for consent from the participant for future research on archived data? Every time we need to do future analyses on the stored data? Already, we had a discussion around informing participants or family physicians. Our international peer review was very adamant that we have to give information back to our participants.

Another question is who determines that it is a clinically significant finding, and do we send it to the family physicians? Do we have to ask the participants whether we can send this information to their family physicians? And in Canada, we have a shortage of family physicians. So does this burden family physicians? Are people going to jump queues to see physicians? There is a whole slew of discussions that are coming up in relation to that. Also, we have a huge issue, as some studies here have, about overarching IRBs vs. local IRBs. And then we have data-linkage issues. Privacy and confidentiality of the linked data. That is a huge issue and we are working with the privacy commissioners within each province to sort out these issues. And, of course, the whole issue of security around public access at CLSA. We are doing some pilot work in our initial phase to look at some of these logistical, ethical, feasibility issues. The phase 2 of the development will focus on development and evaluation of measurement tools required for the CLSA. The proposed launch is in 2008.

Human Subjects Issues Case Study

Speaker: Karin Rhodes

Chicago Conference on Biomarker Collection: Human Subjects Issues Case Study

Karin Rhodes, MD
University of Chicago

I think it's interesting that everyone's gearing up to collect biomarkers as part of population-based surveys. Having done screening for violence and abuse for about five years now in a clinical setting, I'm aware that the complexities and ethical dilemmas that come up in real cases will not fit your protocols.

Case

- 78 y/o AA F Chicago's Southside
- 9th grade education
- Multiple medical problems (PMD 1/month)
  - NIDDM, CHF, HTN, Arthritis on multiple medications
- Widow cared for in her own home by 26 y/o nephew recently released (on electronic surveillance) from CC Jail pending sentencing for property crime
- Younger sister, daughter, and granddaughter in neighborhood visit once week – take her to church

This is an amalgamated case from some of my ER experience. I want to tell you about a 78 year old African-American woman on Chicago's south side. Your interviewers have gone to her house, she has a ninth grade education and multiple medical problems. She sees her private MD about once a month for non-insulin dependent diabetes, congestive heart failure, hypertension, arthritis, and is on multiple medications.

Physical Environment

- Run down neighborhood, home
- Nephew smells of ETOH, cigarettes
- Stuffy overly warm environment
- Dirty dishes, remains of dinner left out
- Fall hazards noted, no grab bars in BR
- Refrigerator w/ minimal food but much beer
- 4/8 prescribed medication bottles empty
- No real privacy during survey

She's a widow, cared for in her own home by a 26-year-old nephew who has recently been released from Cook County jail. When the jail is overcrowded, they frequently release people who aren't a flight risk and sometimes put them on electronic monitoring, also called an 'ankle bracelet.' So this woman's nephew is on house arrest. She also has family in the area that really cares about her. Her younger sister is the mother of the nephew, and her daughter and granddaughter visit her in her home once a week, clean her up, take her to church. So that's the situation.

One issue here is whether or not you can observe some of the physical environment, and some of the very obvious things you see as you go to her house. I wanted to throw that challenge out. It's a very run down neighborhood, people hanging outside of apartment buildings. Her home is rundown and the porch railing is broken. When you go into the house, her nephew opens the door. He smells heavily of alcohol, cigarettes, and possibly of marijuana. One of his friends, who has been hanging out, takes off when you arrive.

It's a very stuffy, overly warm environment. There is no air conditioning during the Chicago summer. There are dirty dishes and the remains of food from the night before left out. You notice a number of fall hazards: Throw rugs that could slip and lack of grab bars in the bathroom. The nephew offers you some ice water. When he opens the refrigerator, you note, there is almost no food in it, albeit there is a lot of beer. When he shows you her medication bottles, four of the eight medications that she's been prescribed are empty, and he explains that he hasn't been able to get out, because of the house arrest. He also hasn't been able to get his aunt to the clinic and she missed her last doctor's appointment. The nephew hovers around as you begin the consent process. You explain that you have to ask the survey questions in private and he goes into the next room. However, his aunt is very hard of hearing and you have to speak fairly loudly, so there's no real privacy during the survey.

Biomarkers by observation

- Obese unkempt woman sitting in chair in dressing gown @ 3pm watching TV
- Pleasant but w/ depressed affect, hard of hearing
- Smells: urine, body odor, baby powder
- Bad breath 2nd poor dental hygiene noted with oral swabs
- Multiple bruises on arms and legs
- Edema lower extremities, shin w/ infected ulcer
- Limited mobility gets up w/ assist (uses walker)

Now let's consider the biomarkers that you're collecting, both for the survey and by observation of the patient and environment. The subject is very obese and not well-cared-for. She's sitting in a chair, with her legs hanging down; they are very edematous (swollen). She hasn't had a recent bath or gotten dressed; it's three p.m., but she's still in her dressing gown, just sitting watching television with the volume turned way up. She's very pleasant, but has a fairly depressed affect, and you have to speak very loudly to get a response from her. However, she seems to understand the consent form. The smells are oppressive in the place. You smell urine, she has a Depends (adult diaper) on, and it hasn't been changed. There's a lot of body odor and baby powder.

When you go to do the oral swab, she has poor oral hygiene and bad breath. It has been a while since her teeth have been brushed. When you take her blood pressure, you can't help but notice multiple bruises on her arms and legs. However, she's supposedly taking a blood thinner, and she says she bruises easily and has had recent falls -- nothing like loss of consciousness or anything. You also notice that her shin has an ulcer that looks like it's an early infection. She has very limited mobility; she usually gets around with a walker. However, she fails the get-up-and-go test and requires a fair amount of assistance even to stand up. But nonetheless, she is alert and oriented and only mildly cognitively impaired, which might be attributed to her ninth-grade education.

You don't feel like she's literate enough to complete the self-administered portion of the survey. Her blood pressure's elevated. Again, she has missed taking some of her blood pressure and heart medications and becomes out of breath with minimal exertion. Her peak flow is about 50, but this is based on very poor effort, so it's difficult to assess if it's accurate. You do your home blood tests and find that her hemoglobin is a little low, which could be a sign of chronic disease or loss of blood from gastrointestinal blood loss. Her blood sugar is elevated at 250, and you follow the protocol and tell her that says she needs to see her doctor as soon as possible. You also tell this to her nephew and he says that she does have an appointment in three or four days. Even though she missed the last one, he's arranged with his mom to help get her there. Each doctor's visit requires arranging for a 'medi-car,' because she's fairly large and requires lifting help as well as transportation. You really are unclear if you can get the urine and vaginal swab on this woman without assistance, and you are hesitant to ask her 26-year-old nephew to help with that. She's very obese, so it's hard to get in there, and you're somewhat confused as to whether or not at this point in the protocol you should actually send her to the bathroom and try and help her or what.

Questions

- Are there ethical issues here? – if so, what are they?
- What, if any, actions are advisable for an interviewer in this situation?
- If actions need to be taken, what should they be? How urgent are they?
- Should this patient be asked about abuse?

Here are the questions that I'm posing to the panel: Are there any ethical issues here, and if so what are they? And what, if any, actions are advisable for the interviewer in this situation, and if there are actions that need to be taken, what should they be? How urgent are they? And should you ask this patient about abuse?

More Questions

- Does it matter if the subject discloses vs. denies abuse, when you are suspicious?
- What cautions should be in place when we consider reporting requirements?
- What further information would be helpful in deciding whether abuse is an issue or the extent of neglect/abuse? How should this be collected?
- What responsibilities do the PI's have in this and similar situations?

Then, furthermore, to put all the questions out in front of you: How important is it that you ask this woman about abuse? If you're suspicious, do your actions change depending on whether or not she discloses abuse? What precautions should be in place when we consider reporting requirements? What further information would be helpful in this case in order to decide whether abuse is an issue and the extent of neglect or abuse? How should this further information be collected? Lastly what responsibilities do the principal investigators have in these and many similar situations?

Savage: I won't get into the research ethical issues because that's not really my training, but when I was looking at this before you gave us more detail, I was thinking 'okay, slip and fall hazards.' We have many worker's comp claims. They're usually in our field organization, and a lot of them are slips and falls of interviewers; not respondents. You mentioned in your example that a porch rail was broken. One of our most expensive worker's comp claims, about two years ago, involved a broken porch rail where an interviewer slipped and fell and did major damage. Now, those are really not unusual, and have nothing to do with the biomarker issue; they are field interviewing issues. So, while they're very alarming, those issues are always there. That's all I'll say about that. The other thing you said that really scares me, though, is that the respondent might not be able to do the self-administered urine and vaginal swab without assistance. To me, there should be no lack of clarity about who collects the sample. I don't know your survey's design. When NORC does a survey, we either know we assist, or we know we do not assist. The reason that matters so much is because when you go to get your risk insurance for the survey, there's a total divergence dependent on that. There's a huge difference in the perceived liability issues. You go down one path vs. the other path. Let's say we go down the path where we've all said no, we will not give any assistance on this particular biomarker. But, then, you know, someone asks for assistance, or maybe you just think, 'Oh, this is never going to happen unless I help them.' It scares me that there might be any lack of clarity. We've got to follow what we said we would do - the protocols - , or we can have claims that aren't covered, and they could be astronomical claims. They could wipe out a company financially. These aren't research issues, and yet, they are, because they could wipe out a company. Those are my two biggest issues.

Rhodes: Do you allow the family to assist?

Savage: Well, that's another issue, because suppose there are many things if the family assists. You don't know those relationships, and you don't know people's comfort levels. There're all kinds of things, all kinds of things. Those are just the first two. There are many others.

McPhatter: This case is very, very interesting, and actually reminds me more of claimants that I dealt with as a disability examiner, adjudicating cases of Social Security and Medicaid, and some of their issues, than actual delivering protocols for test-result counseling in research-based studies. Something like this -- I don't know the biomarkers -- would she be calling a service to get answers to the questions? We already knew what her medical status was, and I don't know how involved a service like ours would be in a particular case like this. But to me, on an ethical side, this raises a lot of issues in access to care. Some of the main things that I saw were the limitations that she had, either as a result of not having support services, medical assistance and things like that, or maybe the home environment with the existing caretaking in place, not being of benefit to her. In either the survey or this study, I don't know if it would be appropriate for you to make those kinds of determinations, but I could see a person like this being eligible for a lot more support services than she was obviously getting. I don't know if the field interviewers are authorized to give other information, and make referrals for the things outside of what they're there to study and the information that they're there to collect. That might be something to consider.

I don't think it's very uncommon to see cases like this. It's heartfelt, but it's very, very prevalent in our society, especially with older populations that might not know which support services are available. They just haven't accessed them because no one has told them. Or their family members have been limited. In this case, he's recently out of prison, with limited transportation, cognitive impairments with alcohol and other mental issues, perhaps, which really don't make him a good person to be advocating for any health resources for her. And so to me, that's more of the issue. More issues than we can focus our attention on.

Unidentified Speaker: Can you imagine a situation where, say, a field professional identifies, they don't know exactly what the problem is, it just seems like not a good situation? And refers to ASHA through the auspices of this particular study, the person to call for the kinds of referrals that you mentioned. Is that something that ASHA's ever done, or can you imagine ASHA being involved with that sort of counseling?

McPhatter: The only way we would do that is if it were part of the protocol for a particular study. We have a nationwide database for information referrals, health care providers, support services, and things like that. But from the general perspective of someone not affiliated with the research study calling us, they would be calling for specific information. Either immunization information, STD and HIV, herpes, HPV, not because there were some care issues at home. That would be separate. If we were involved in a study similar to this, and a participant did call and either mentioned to us or we had some access to this type of information, then we would definitely make the appropriate referrals for medical attention and other caretaking resources. But of course, it would be an external referral. We really wouldn't do a lot more.

Unidentified Speaker: When people do call in for results, how often do the people responding have to go into these peripheral issues?

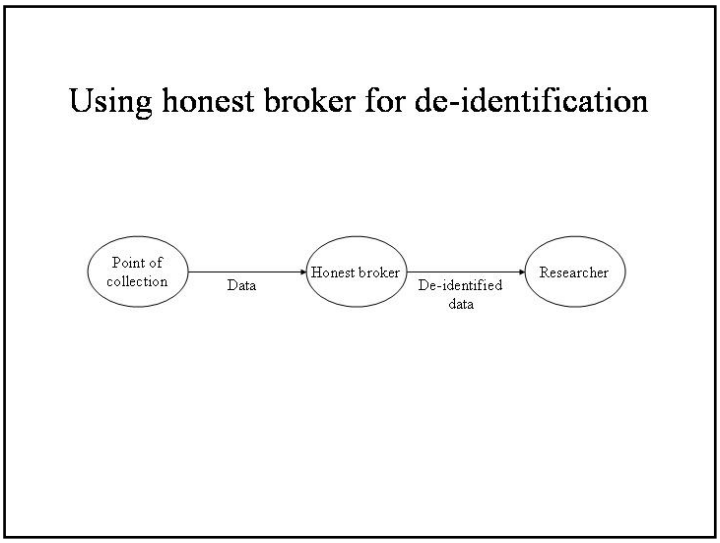
McPhatter: Again, that varies. I would say that most of our studies have been targeted to a set population. Then, during the screening process, there is some pre-test counseling that goes on, so they really are informed of what our limitations are in information. They really wouldn't necessarily bring that up. Now, a lot of times in partner communication, whenever we give results, and I'm going to speak of this from an STD perspective, we do the emotional and cognitive integration of the information that we share, to see how well they're able to understand the information, and what other type of support services they may need. That's part of the protocols that we do, regardless of what they tell us. In a case where we specifically notice that there were some other things that they would need, then we would make those referrals. Get them to a place where, you know, they would feel empowered enough to see that seeking additional things are things that they should do. We wouldn't make a recommendation of them doing that necessarily, unless they asked, and that would be a referral. We would be globally assessing the whole situation and getting them to really cognitively look at it and make decisions that they feel are in their best interests.

Honest Brokers:
What are they, and do we need them?

Honest Brokers:
What are they, and do we need them?

Speaker: Phil Schumm

I'd like to give a little bit of background, at least, on the way that I came to this issue, and the way I think about it. I certainly am not an expert on it, but through our work in the Biostatistics Laboratory in the Department of Health Studies, we've had a number of multi-center studies that have been forced to deal with this issue, and so we've acquired some experience with it as a result. One of the most important points I'd like to make is that there are at least two different ways of thinking about what you actually accomplish with an honest broker.



This is the first example. What we have, essentially, is the point at which the data are collected, and I purposefully have labeled that 'point of collection' rather than 'study subject.' It could be lots of different things, but the idea is that you can't back up any further in the chain than that. And the idea in this case is that the honest broker is an intermediary that serves in-between the ultimate researcher and where the data are collected, so that all of the data, essentially, are passed through that individual or organization. The data are de-identified, and then passed to the researcher.

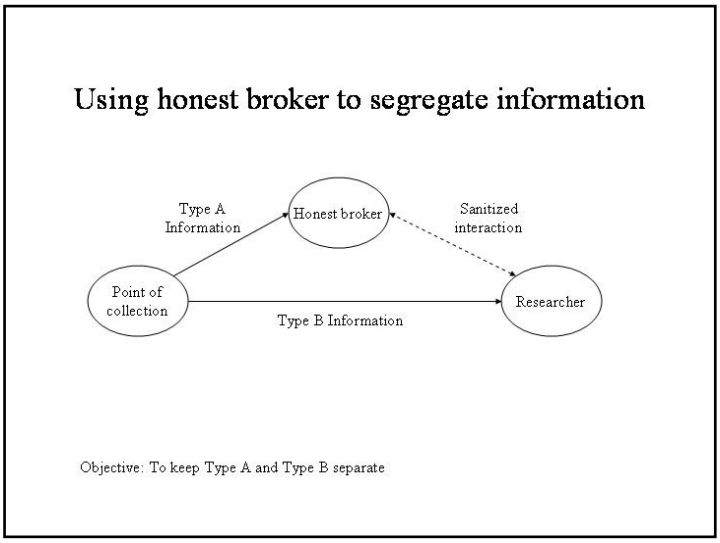
Now, a couple of things to note about this: one of them is that I would argue that really, of course, all this is doing is just moving the burden of confidentiality from the researcher to another party. That's all it's doing. As a researcher myself, I'm not sure that, if I were asked to give data, that I would trust anybody you could stick in here any more than I would trust the researcher. I think there may be situations where people might trust somebody. A good example would be the recent study of priest/child sexual abuse, by the John Jay College of Criminal Justice. In that case, the point of collection was really the individual dioceses who were asked to submit lists of priests who had been accused, and they were asked to submit the date of birth and initials of those priests, so that, ultimately, priests who had been accused in multiple dioceses could be identified, and you could de-duplicate the data: you wouldn't count them twice. In that particular case, the data were submitted from the diocese initially to Deloitte and Touche, who de-duplicated the data and then passed them on to the researchers. That, I think, was an important way of making the church feel a little safer about it. I think, in part, possibly because they were distrustful of some of the motivations of the researchers. In my experience, however, certainly in the kind of research that we do, most of the study subjects, once you've gotten in the door at all and they've agreed to talk with you, usually believe that you're about doing research. That's not really their concern.

Just two other things to note here. One is that if you're doing any kind of a longitudinal study, where you have to go back to the subjects, or a study that requires a lot of long-term follow-up, so even certain

kinds of genetic/environment interaction studies, is very problematic, because it means the honest broker is on the hook for the duration of the whole study. That's something that I'll talk about in just a moment. Certainly, those of you who are clinician researchers will notice this model doesn't really apply if the researcher's actually the one who's collecting the data.

Unidentified Speaker: I have one comment also, that in the other place where the honest broker's been used is in clinical trials, so that the researcher doesn't know the outcome of an intervention in a way that could affect the trial. In that cause, you need to have data monitored during the trials so that, say, if the effect of one drug was very negative, the trial could be stopped.

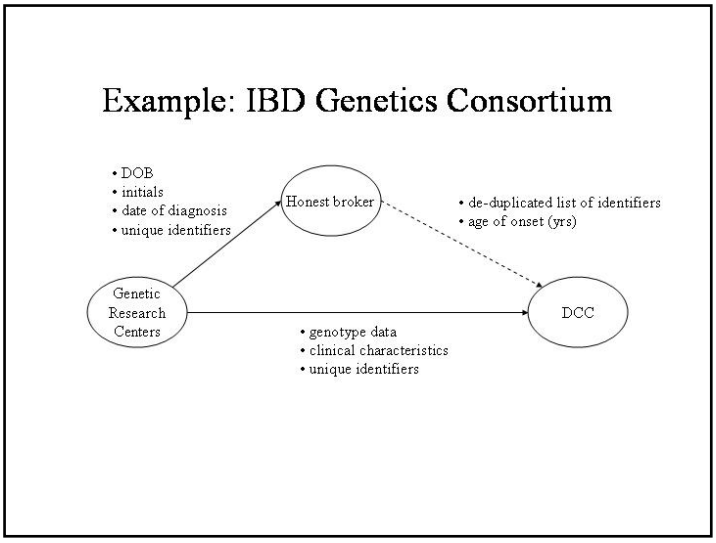
Schumm: We do that all the time in a data-safety monitoring board, and so that's not so much an issue. Usually in that case, there's an agreement between the data safety monitoring board and the researchers that, during that process, the data go directly from the data coordinating center to the data safety monitoring board, and the researchers simply agree that we won't actually take receipt of the data until the end of the trial. It's a really good point, but of course, once the trial's done, that issue is no longer an issue.



This is another way to think about what you can accomplish with an honest broker, very different from the first. And this is actually sort of the way I prefer to think of it. In this case, you have again the point of collection, and another thing that particularly comes up in this model, that I haven't really shown you here, is that the point of collection can actually be multiple different entities, or there can be people involved in the pathway between the point of collection and the honest broker: for example, a laboratory which assays biological samples.

In this case, the idea here is that information of type A go from the point of collection to the honest broker, and information of type B go directly to the researcher. The whole idea here, what you're trying to accomplish, is that the information of type A should never exist together with the information of type B. Often, what we mean by type A information are identifiers about the person. But that doesn't always have to be the case, and I'll give you a little more complicated example of where it's not as simple as that. But that's essentially what's accomplished here.

Now, of course, if the honest broker just held on to those information and never talked to the researcher, then you didn't have to collect information of Type A to start off with. So the whole idea here is that this is information that becomes important either during the remainder of the conduct of the study, or during the analysis, or in perpetuity further on, and as a result, the honest broker and the researcher frequently need to communicate with each other. So the idea is to set up some kind of what I've called 'sanitized interaction,' which allows them to do that in such a way that, again, this type A information and type B information never come in contact with each other. That's not a way that I've seen before to describe what you're trying to accomplish, and, at least for me, it's a good way to think about it.



Let me give you a couple of examples then, quickly. One of the things that I'm involved in is a Data Coordinating Center for a genetics consortium funded by NIH to study the genetic basis of IBD. In this particular case the consortium does many things, but here's one of the first projects that we got involved in. The Consortium essentially consists of the Data Coordinating Center in Chicago, and then six different Genetic Research Centers (GRCs) across the country and in Canada. One of the first things that happened when the consortium began was we realized that a lot of the centers themselves had collected a tremendous amount of information that would be awfully useful for a genome-wide screen to start the project off. Nobody had been able to analyze their data together, and so in determining the future scientific directions the consortium should take, it turned out that would be a useful thing to do.

The problem is that these data had been collected by the GRCs over fifteen to twenty years in some cases, and so it was very unclear what the people who had given their blood at that point had consented to. Some people had said they would share their sample with other researchers in the future, but in some cases, there was some ambiguity there. So what we needed to do was to figure out how they could send us their data. The problem was this: we needed, obviously, the genotype data, and then clinical characteristics to actually do the genome-wide screen. But the problem is, as it turns out (and I was actually very surprised – I didn't believe this at first), but as it turns out, particularly for diseases like IDB, it's not unusual for people to go across the country and become involved in these studies multiple times. Particularly in a genetic study, where you have different people in the family being involved. The way the samples frequently are collected for family members, of course, is that you have the proband come into the clinic, and then, for the family members, you actually mail a kit to them, which they take to their physician. One of the things that, as a statistician, I said, was that this will happen in a handful of cases, but it's not going to affect our analysis. I wouldn't worry about it. But the geneticists were very worried that, when they submitted their paper, they would be able to say that they de-duplicated the data. To do that, we needed date of birth and initials for the individuals.

We also wanted the clinical characteristics. We were particularly interested in what the age was at diagnosis. And so, while people can, of course, submit age at diagnosis, again, as a statistician, I'd have lots of people bring these spreadsheets with date of birth, date of presenting to the clinic, or date of diagnosis, and then age of diagnosis. You'd be amazed at how many times age at diagnosis is not consistent with the first two pieces of information. So we said that we'd prefer to calculate that ourselves. The problem is that the IRBs at several of the genetic research centers did not feel comfortable with genotype data being put together with identifiers and sent to us. Some did, but some didn't. And so, to make this happen, what we did was, we enlisted the services of a statistician honest broker, and set up, there's actually more complexity to this – as they submitted this data, it was also encrypted and that sort of thing-- but they submitted these data to the honest broker, who was then able to use them to do two things. Basically, to de-duplicate the data, and then also to compute age at diagnosis. And you notice that each of the GRCs maintain their own unique identifiers. They sent those identifiers together to the honest broker. The genotype and clinical data came directly to us, and what the honest broker was then able to forward to us was a de-duplicated list of these unique identifiers, together with age of onset.

Unidentified Speaker: But you did end up getting unique identifiers directly from the GRC?

Schumm: And by unique identifier here, I don't mean Protected Health Information (PHI), I simply mean some piece of information that's unique to each person in the study. Those would be, in this case, the pedigree and individual numbers that they had assigned. But they don't mean anything to us. We can't use them to find anybody.

Unidentified Speaker: But that unique identifier went to the honest broker and also came back, in a de-duplicated form?

Schumm: Exactly. That was critical, that was what basically allowed us to make sense of the information that he was giving.

[audio unclear]

Unidentified Speaker: So you need some non-HIPAA unique identifier, like a pedigree.

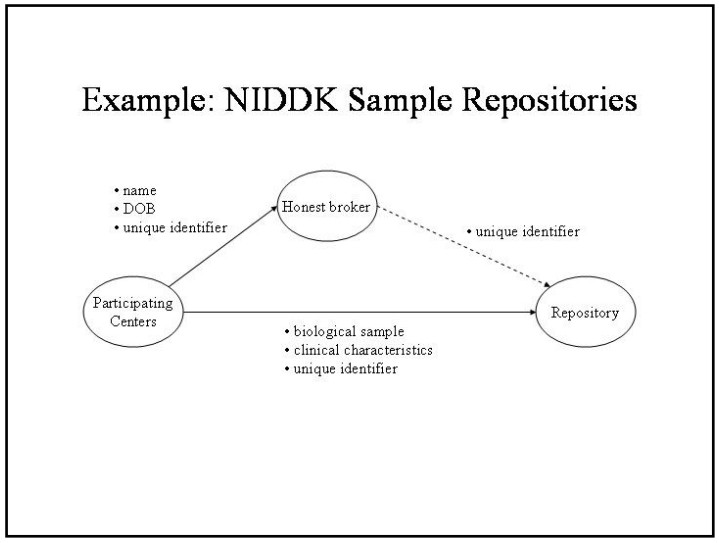
Unidentified Speaker: Well, age in years is not PHI, age is okay.

Unidentified Speaker: But age alone is not a unique identifier.

Unidentified Speaker: Right, exactly, but it's some sort of-

Schumm: We basically stipulated that some people use parts of medical history numbers and things like that in their medical I.D., and we stipulated very carefully that it couldn't include those sorts of things. All of those things first went to the honest broker in encrypted form. He unencrypted them and made sure that these were acceptable to send, and it was only at that point that they actually sent these things to us.

Here's another example. The NIDDK has been setting up a number of sample repositories now -- they're in the initial stages of it. In this case, the repositories are to collect biological samples, and also clinical characteristics. These would be samples that then would be stored for genotyping in a variety of analyses in the future. In this particular case, their issue is that they were taking very seriously the stipulation that people who participate in the study, at all times in the future, be able to withdraw their participation if they should decide to. Now, of course, I pointed out that most people after five years are going to forget that they were ever included in this, but needless to say, they took this very seriously. This is another good example of where an honest broker could come into play, although he or she would be doing something slightly different.



In this case, the participating centers, or probably a data coordinating center, stuck right in the middle here, would generate unique identifiers, and then the name, date of birth, and identifiers would be sent and stored by an honest broker. The rest of the data identified only by this sort of anonymous identifier, would come and be stored in the repository, and then at any time in the future, if a person walked into the center where they actually gave their sample and said, 'You know what? I want to be removed from your list,' they could simply send the name and date of birth to the honest broker, who could then just forward the unique identifiers of the persons who should be removed from the repository. The issue of how you store samples is something that hasn't come up here, but I mentioned it only because, in the case of NIDDK, there are usually some very good sub-contractors, who have sprung up. They are usually located at universities to do this kind of storage. It's very important. Being able to interface with an honest broker and a data coordinating center effectively is a non-trivial issue.

Another non-trivial issue that we need to think about in our case, a survey context, is how samples are stored. If they're stored in certain ways so that the repository can use robotics to get at them, when we go back to our samples and we specify either a sub-sample that we want to do an assay on, or we design that sub-sample ourselves, or we want to lay out assays in a certain optimal way to look at certain things, if the repository doesn't have the samples stored in a certain way, so that you can use them with robotics, that's going to be impossible. The reason I mention that simply is that with this whole issue of setting up an honest broker system, together with places who might be storing and processing your samples, you have to think about it holistically. You can't put the pieces together separately and expect them to work together. That's been one of our biggest challenges.

Unidentified Speaker: And the other issue is that, potentially, you could have a lot of extra biases that you have to analyze, if you do have people dropping out in a systematic way. Certain ethnic groups that might become more suspicious after other things come up with it.

Schumm: What that means, basically, is that we have to determine who's been dropped out.

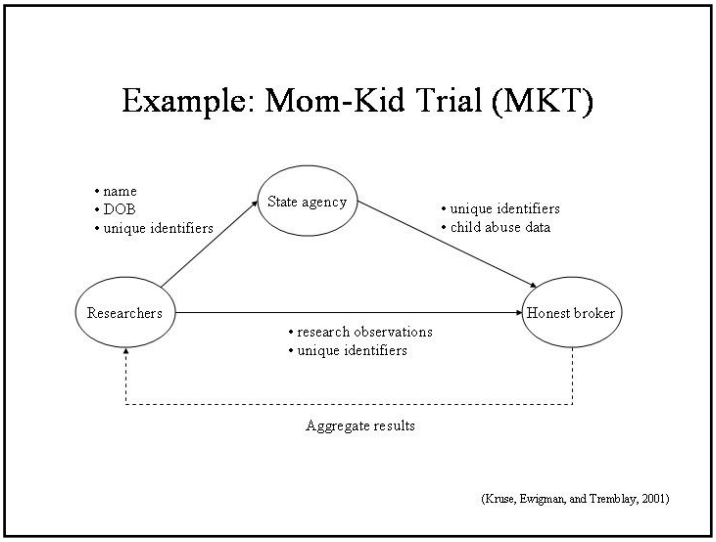
Unidentified Speaker: What are the requirements for becoming or being an honest broker? I mean, what does the honest broker have to do?

Schumm: Well, there are no formal ones. I have a few thoughts on the matter, but part of the reason I was mentioning this is one of the things that we decided in working with NIDDK is that there do need to be some sort of sub-contractors who do this thing and do it well, because for studies that need to go off into the future, you can't be dependent on an individual person. In the case of the example that I showed you before, it's actually a statistician who's down the hall from me. He doesn't have the virtue of being in Canada, but he does have the virtue of being very trustworthy.

Unidentified Speaker: About the repository, does that just pull data for the P.I.'s in the survey? Does it release data to the public? That is a concern.

Schumm: The whole point of it is to be a public repository, exactly. It's one thing to send biological samples off to people, but for a lot of the studies, they're not too interesting without some of the clinical data as well.

Here's a slightly more complicated example. That's all I'm going to say, but, I think, this is useful because, you know, the idea is to be very creative here. This is an example that some of you may be very familiar with, from something called the Mom-Kid trial, done at the University of Missouri in their child protection research unit. This is a system that they developed initially, and then subsequently used this in something called the Mom-Kid trial.



In this case, the researchers are psychologists who are collecting, in a clinical setting, observations from a parent and a child. They are particularly interested in whether those observations can predict the likelihood that the child will be abused in the future. The outcome, essentially, is something that's collected by a state agency – in this case Child Protective Services, in Missouri – and those are essentially lists of complaints both about potential abuse of a child who had been included in the study, or of a mother who had been included in the study of an additional child. They had all of their observational data, and essentially wanted to put it together with those outcomes. Those are the two pieces of information.

The problem here is that the state agency couldn't release those data in a way that the outcomes could actually be put together with an individual. And of course the clinicians couldn't release any of their medical information in a way that could be identified by anyone else. The way it was solved in this case was that the researchers essentially created a list of unique identifiers, and sent two data sets. One to the state agency that included name, date of birth, and then this anonymous I.D. The state agency essentially identified those with complaints they had received over a specific period, stripped off name and date of birth, and then forwarded what was essentially the outcome data together with that anonymous I.D. to an honest broker. The researchers then also sent the honest broker their research observations, essentially their covariates in this study, together with the same identifier that they had sent the state agency. At that point, the honest broker could put together the output data with the covariates here. The honest broker was the one who actually did a set of analyses stipulated by the researchers, and then the results of those analyses were passed back. It's something that would be difficult for us to imagine doing, but there would be variations on that that you could actually imagine implementing.

Unidentified Speaker: Dr. Whitman, who did that study, is at the University of Chicago.

Schumm: Yes, you can actually read more about this. I forget the journal that they published that in, but yes.

Final Thoughts

- Purpose: de-identification versus information segregation
- Do not underestimate:
  - Burden on honest broker
  - Effects on research process
- Should be used only to overcome otherwise unsolvable problem

So just a couple of final things. I think that you need to be clear about the purpose of the honest broker, what you're trying to accomplish with it. I get a fair number of calls from people who say they've heard about this, it's something they're interested in, and the first question that I ask is what they are trying to accomplish with it. They know it's a way of improving the human subjects' protection, but at that point haven't identified a real purpose. The reason that that's important is for two reasons: for one, don't underestimate the burden that this places on the organization you choose to be an honest broker. And that burden comes from a number of things. There's a big data management task here, and we talk about this, and I'm sure you're obviously intimately familiar with this. I guard my data pretty jealously, and I don't trust anybody else's manipulations on it. So imagine sending your data now out to somebody else. You're asking them to frequently do pretty complicated things that the results of your analysis will depend on. This is one of the reasons why, as statisticians, we like to send them to statisticians. There're obviously lots of other people who could do that, but it takes a lot of time. In cases where we are collecting the data prospectively, we can have a fair amount of control over the quality with which, and the standardization with which, they're collected. But when these are data that have been collected all over the map, putting that together is an absolute disaster. And that gets me to my second thing. And so the honest broker told me to 'make sure to tell everybody I'm a disgruntled honest broker.' Over the past couple of weeks, they've decided since the initial study we did with those data to go back and actually look at it for some other things. And that's required me sitting down with him and going through the data. Of course, that's a rather strained and funny interaction, because I'm sitting on one side of the computer, knowing that I can't, in principle, look at the information that he has, but essentially telling him what I want him to do in the columns and the data. It's a big issue. And if he were hit by a bus tomorrow, we'd be in trouble as well. So issues of redundancy and so forth come up. Oh, I mentioned sort of the notion that we could possibly sub-contract with an organization, but it would obviously have to be somebody who had research experience. The other thing is that I wouldn't underestimate the effects that this has on the research process, and those are several: one is delays. In our particular case, we had some data but before he de-duplicated the data, he wanted to just do it once. It was a big laborious thing and he wasn't being paid, but one center had delayed in sending him their information. That meant that we couldn't start with our analyses at all until he got that. And so delays are an issue. Another is that whenever you create a barrier between

two pieces of information that you've collected as part of a research project, things that seem very reasonable and normal to do, you can't do anymore. One problem again with integrating data from people who you've collected a long time ago is that you get a submission and then, weeks later, they say, 'oh we forgot some cases and we have some areas that we want to resubmit.' That case is pretty much a disaster, and in fact, I would not do that again, but things that I want to do, like internal consistency checking, to make sure they're sending me the same data they sent me before. If I had date of birth and initials, that would be something I would use in a completely benign sort of way, but it's something that I can't, because we've written into all of our IRB agreements that we won't, we can't. Then the last thing is future work. You jeopardize your ability to put this together in interesting ways with third sources of data in the future. So, for all of those reasons, the one message I hope to communicate is that in each of the examples that I showed you, an honest broker was being used essentially to solve a problem which otherwise couldn't be solved. There was no other way to move forward with the research. And that, I think, is the most useful way to think about what you do with an honest broker. Not something that you just slap on because it's nice to have some additional protection and there's some sort of vague feeling that others are doing it, and it makes things look good. If it's actually a way to hit a particular nail, then that's great. But if it's not, it's probably not something that everybody needs to be getting involved with.

Unidentified Speaker: Phil, I want to suggest that you think about writing that into a paper. There are no publications about that. There is really such a lack of information about this, and even trying to educate people we wanted to talk about it today was difficult with the limited information.

Unidentified Speaker: As far as the disgruntlement of the honest broker, can that be addressed through budget issues? I mean, even forwarding your future projects, is this something that could be incorporated into the budget of the research project? And would that address the concerns of the honest broker and make that person feel like what they're doing is valued and is sort of accounted for?

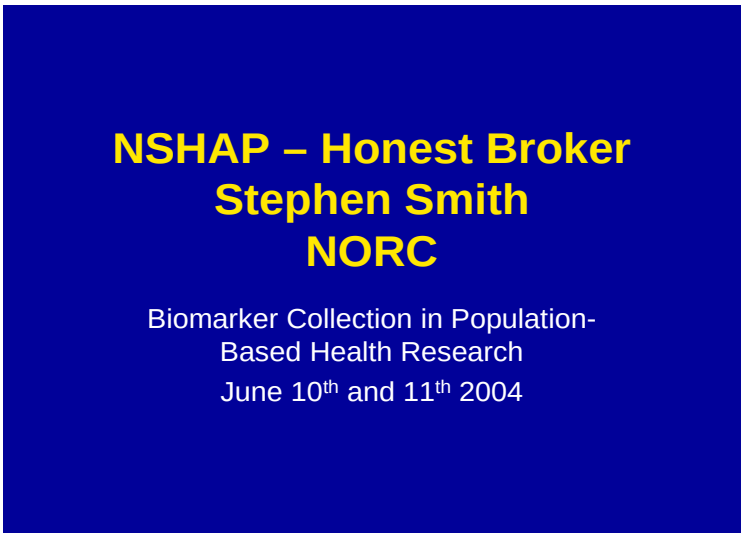
Schumm: In our case, it couldn't have been dealt with that way, because the person was doing it partly to be helpful and because we had entrusted him, but not essentially for the money. It was a person who had a fixed amount of time to spend, and more money wouldn't have bought more time on his plate. I do think, however, that that speaks to the issue of possible subcontracting here. The subcontractors with the repositories we've been working with are not people coming from industry. They're primarily people with academic research labs who have the capacity, who have some interest in doing this work well, and so they actually know how to do research, and they're much easier to interface with. You can imagine a data consulting firm, Andersen, coming in, and, of course, you would pay them to do this. I would feel much more comfortable, however, if there were a center that had experience as a data coordinating center who, for whatever reason, saw this as something that would be, you know, useful for them to do.

Mullan Harris: I just actually wanted to make a comment about the honest broker, because some people would see it as removing the issues about confidentiality from the researcher. I think that, at least, that's not how we view it, because the risks of deductive disclosure are huge. I think with most – not even Add Health, because it's so contextual – but most of the studies, it doesn't take many variables to identify. So we impose a lot of additional restrictions in terms of disseminating data. Our researchers have to think all the time. They need to spell out their security plans, depending on what level of data they have. But you know, we really sensitize them to the fact that they are just as vulnerable to confidentiality breaches through deductive disclosure, so the honest broker isn't the solution to issues about deductive disclosure.

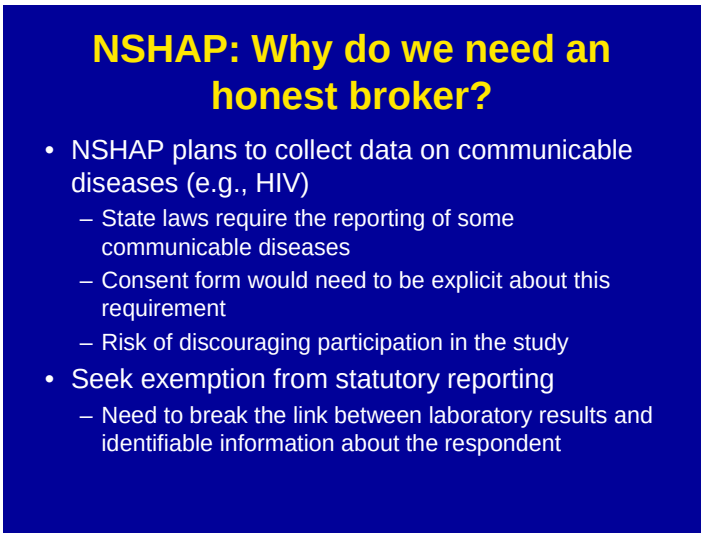
Schumm: That's absolutely right. This was more organized around clinical studies, where you have less of a chance for that. Obviously, with a survey, as long as you've got the data together, you've essentially got the possibility of that, so it's really a different issue. You can still, however, imagine the third example I showed, where if there were certain things that were very sensitive, but that were critical for you to collect if it was part of your scientific goal, that you could create a way of still getting those analyses done without ever putting those together with the rest of the main bulk of the data, but you're absolutely right, that's true.

NSHAP – Honest Broker

Speaker: Stephen Smith

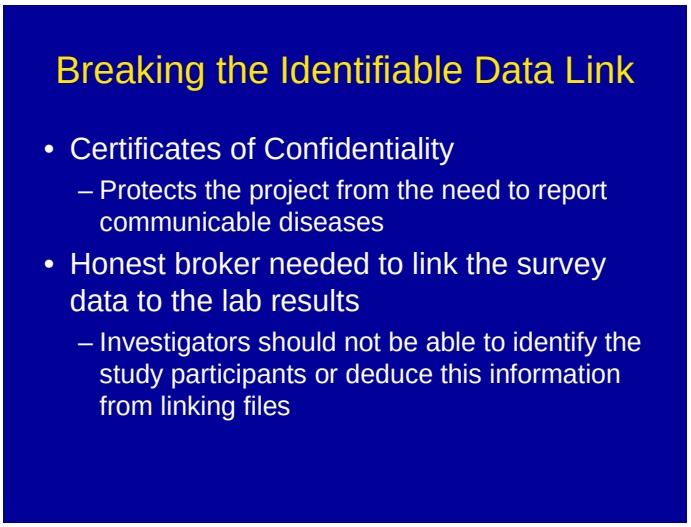


I'm Stephen Smith from NORC. In the case of NSHAP, and this is generalizable to other studies collecting communicable diseases, we're going to be collecting HIV data. State laws require the reporting of some communicable diseases. That's a given. Now, if you didn't want to do anything else, the consent form would need to be explicit about that requirement: a respondent would have to sign the consent form that states that the results of the HIV test is reportable. So out of concerns of discouraging participation in the study, we need to go to some lengths to avoid having to report these results to the state. You need to seek an exemption from that statutory reporting requirement, and the way to do that, just as Phil described, particularly for communicable diseases such as HIV, is you need to break the link between identifiable data and the HIV status, the outcome. Two ways to accomplish that goal is (1) the certificate of confidentiality, which has been mentioned. You do have to jump through a number of hoops to get one of those, and I learned yesterday that the certificates have never have been tested in the courts. In principle, it does protect the project from having identifiable data subpoenaed, so that's often a desirable certificate to get, but it does require a lot of administrative work.



And then there's (2) the honest broker needed to link the survey data to the lab results, which is what Phil described. There's one general principle here, the investigators – and personally, I'm struggling with how to define what "investigators" means in this context – should not be able to identify the study participants, nor deduce who they are from linking various files. That becomes very challenging. I know in Add Health's case, and certainly in our case when we get further down the road, that it's going to be a real issue.

Mahay: Just a point of clarification, you said that an honest broker is required in order to get the certificate of confidentiality…?



Smith: No, no.

Unidentified Speaker: We would have to get the certificate of confidentiality for the project.

Mahay: Right, but in order to get that, do you need to have an honest broker system?

Unidentified Speaker: No, no.

Smith: It's a Catch-22, I know. I've tied myself in knots trying to work this one out.

Mahay: If you have the certificate of confidentiality, that means you don't have to report data?

Unidentified Speaker: First of all, it's not been tested, so-

Mahay: So it's just sort of a failsafe issue?

Unidentified Speaker: No, the other issue is that, separate from reporting, there's the potential for subpoenability of records. So, in our idea, an honest broker who's outside of the United States and Canada and is the only one holding those data is much more difficult to be subpoenaed, say, by a husband who is in a divorce case and wants to know if the wife is cheating on him or something.

Mahay: But the certificate of confidentiality-

Unidentified Speaker: We don't know, it's never been tested.

Waite: Which means nobody's tried.

Unidentified Speaker: Well, we don't know that nobody's tried. It may not have got far enough.

Waite: I think we're being paranoid about this stuff.

Mahay: My question was, is the honest broker system the failsafe, so if the-

Waite: Yes. We think it's a failsafe. It hasn't been tested either.

Mahay: But in case the certificate of confidentiality doesn't work-

Waite: It's the bomb shelter we're building in the backyard.

Unidentified Speaker: I think it's really a system that's set up to increase your pledge of confidentiality, as long as you don't have that locating information together with the data in one place, it just reduces that risk.

Unidentified Speaker: I have certificates of confidentiality for almost everything I do, and when I first started to apply for it – and you applied for it through your funder – they just gave it to me. Now, there's a clause that the university has to sign that they will defend it before NIMH will give me my certificate. But they did. The university did sign it, they would defend it. And the bottom line, you have to be willing to go to jail to defend it, just like if you were a reporter and taking confidential information.

Laumann: Some years ago, the university commissioned a review of the law, and at that time, we explored various ways of trying to corrupt data to make it less penal-proof. For example, the idea was to [audio unclear] a certain random variable so that we could not say that we knew for a fact that this individual was the person in question. He looked at that, and as a lawyer, he was saying, 'let's say you had a situation where the person was trying to get certain remedies, or having been infected by knowing an HIV person, who she happened to know participated in the survey. She's looking for this particular source, and even though we would say, there's a five percent error in the identification of who that is, that still would be data that would be better than anything else. Clearly, then, he knew that he had HIV at the time that he was having relations with her.' We then went on to ask the question about putting this in another place, the honest broker, Canada, wherever, and the problem is [audio unclear]

Unidentified Speaker: There is somebody who has information that would be happy [audio unclear], and if you have explicitly gone to corrupt the capacity of the state court to have the remedy; that is, we knowingly put this stuff somewhere else so that they can't get hold of it, learned contempt [audio unclear] or Switzerland, or an offshore island, or underground. It will be seen transparently as an effort to evade the consequence of the state court's capacity of the right to review this.

Unidentified Speaker: That exact issue came up in the San Francisco men's study looking at HIV, and the investigator, and Berkeley, were willing to turn it over. They said 'Well, we're being subpoenaed; I guess we have to turn them over.' It was protected by a promise of confidentiality, and the investigator went on the television and said I will go to jail, I will not turn it over, and they didn't actually acquire that data.

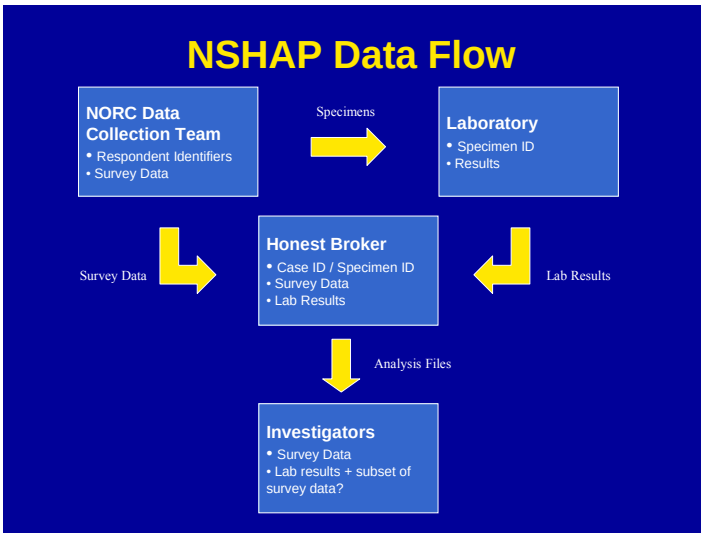
Unidentified Speaker: [audio unclear] – calculated the probabilities that some known person who happens to have HIV, I mean, the chances of HIV are what, 1 in 1,000?

Waite: Not in our sample.

Unidentified Speaker: In our sample, it's probably about 1 in 1,000.

Unidentified Speaker: The whole idea of a known person who has a need to know this information, I mean. It's only just more cooking up of anxiety.

Smith: `I think you're right, everything needs to be kept in perspective. This is our data collection. You can apply to any data collection, it's just describing NSHAP and our current set of assumptions, and how we're going to work this. Data collection essentially means both my research team and the interviewers, since we're sharing data. It's seamless. We're all really in the same box in this regard. We have respondent identifiers, we need them. We're going to be collecting survey data in the interview. And I have to say, this diagram is very much simplified - my apologies. There's a much more complicated version that I make the investigators suffer occasionally. During the course of our study, we are going to be collecting specimens. They're shipped to the lab. The specimens will contain a specimen I.D., because they need some I.D. on them, and then the lab produces some results. Then, the lab results clearly come back to – in our case, we're thinking of using an honest broker. The survey data also makes its way to the honest broker, and at this stage, the honest broker is able to link both the lab results and the case I.D. together. They'll have the survey data and the lab results. At this stage, I won't know the lab results. Only the honest broker does. And then we get into this really tricky issue, which has been discussed by the group today, which is releasing analysis files to the investigators. Clearly, there's a need for something. In NSHAP, we've got the 90-minute instrument. It's an incredibly rich data source, though there's clearly a real strong desire here to have the complete survey data set in one set. Then there's this other knotty issue of lab results, and possibly linking to some aggregate or subset of data so that the investigators in this model won't know the identification of any respondents, but they do have enough data to be able to conduct meaningful analyses.



Waite: But there really only need to be lab results on things that are reportable, which we have already decided wouldn't get us out of this anyway. It would only be the sexually transmitted diseases.

Smith: That's correct, sorry. I should have clarified.

Waite: Not blood pressure, not hearing-

Smith: Not blood pressure, nor any of those, this is very much talking about the reportable diseases. This is a real challenge here, because obviously

investigators and researchers don't really want to over restrict themselves in this set here, and then there's this issue that I'm wrestling with: if this is too long a string of variables, I will, in principle, be able to link the identifiers back to this data set if they gave me too much information here. And that's a knotty issue in terms of how to make that happen and the appropriate protection.

Waite: What about creating measures, outcome measures, that obscured some of this? So I don't really want to know what their tighter level was on some of this, I just want to know, did they have one of a class of diseases? You might be able to create outcome variables that mask enough of this that you could have those on the survey data and only have the detailed information on…

Unidentified Speaker: Except for the most sensitive ones, it's either positive or negative.

Waite: For the HIV-only analyses. But if you only want to know, 'did they have sexually transmitted diseases,' there are certain classes, there might be a way to have an intermediate.

Unidentified Speaker: What if you had a lawyer on the other side, looking through the information. They could go to any pathophysiology book or anything like that, and look at what you do have in your data set….

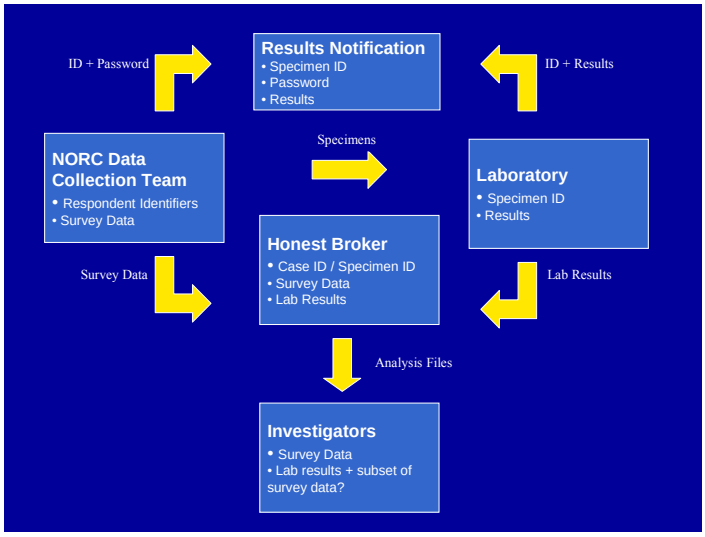
Waite: As far as I know, it hasn't happened. We're worried about all these things that have never happened, and as Stacy points out, if I'm going to be the point person, I'm going to be the one who goes to jail.

Unidentified Speaker: I think there's a bit of paranoia here. Statistics Canada, which is an agency in Canada that does lots of national surveys, they have this paranoia, and they collect a lot of tough information, and they spend a lot of money to be able to, because of these restrictions. People question, what's the point of spending all that money, when you also have so many carriers? And as a result of that, in Canada, it's almost all universities that have research data centers where Statistics Canada or large population-based studies can release their data on an individual level. But you basically have to be sworn in before you analyze that data. With all those resource data centers, you're sworn in, you can't get data. You can spend time, you can look in computers, you go there and do your analysis and come out with your results. It doesn't make life easy, but to me it seems much better than some of the honest broker ideas that have been thrown around here. And you can actually send your own analysts who are experts in data analysis, and they have access to the data. On the other hand, people who are going to violate are going to figure out ways to violate their confidentiality anyways. They're also making a difference between de-identified data and anonymized data. Our ethicists are actually more worried about what we mean by de-identifiable data than by anonymized data.

Unidentified Speaker: I just had two quick practical issues. One is that we have a research center like that at Duke, which is the Triangle Research Data Center. It's for sensitive census analyses. I just wanted to raise the possibility of those folks becoming honest brokers. Second, practically speaking, if you're willing to

make your honest brokers do a lot of work, if it's something like these research data centers, couldn't they hold the data and you get all the survey data, but for anything that's potentially sensitive, you have to put in a request for a specific subset of the data? That gives you a lot of flexibility, if a particular investigator wants to have this outcome with these covariates. Still, that's a lot more work for the honest brokers.

Smith: Yes, I was going to talk about that. Let me just move on, because I've only got two more slides.



So that's the model that I just described. The last piece of the puzzle, particularly in our case, is results notification, which is another tricky issue where you have to have the breakage of the links. In our case, very simplistically, we have an I.D. and a password that the respondent provides that we lose from the laptop immediately after we've captured it. Then the lab can send the I.D. and the results to the results notification service. That's the last piece in the puzzle where there's no link, they don't know the identifiers, etc. Coming back to your good point, which is that this does capture the wrestling match with the honest broker, but it doesn't do away with the much bigger issue that we are wrestling with what data I can give investigators. Then there's data release, and, like public release files, that becomes a very knotty issue, and a complex one, in terms of the data use agreements, and what variables we make available, what geographic identifiers we can include, etc.

Job Vacancy: Honest Broker

- Independent from research team
- Ethical professional
- Data processing and manipulation skills
- Secure environment
- Ability to minimize risk of deductive disclosure from linking of multiple data files
- Time commitment
- Cost

Here is the last slide. It's the job vacancy for an honest broker in the job columns, and I should have no disgruntled honest brokers allowed on this one, but it's really capturing, just in a simplistic way, that it has to be truly independent from the research team. It can't be someone whose arm you could twist to get them to give you the data. That really lends itself to the ethical professional here. It can't just be someone that you really trust, that can feel comfortable with taking on that responsibility. Just as [audio unclear] indicated, it has to be someone with fairly considerable data manipulation and computing skills, to be able to bring this together and

manipulate it, and for the investigators to feel comfortable with that person doing it. It needs to be in a secure environment, both physically in terms of building and in terms of computer security, to make sure that there's no risk of someone breaking into their system. They need to be very savvy about that. This ability to make sure that you don't release a subset of data that allows deductive disclosure is a science in itself, and one I know that the census wrestles with in terms of their release. Time commitment, as [audio unclear] indicated, it's a big investment, not just for the course of the study, but post-study, pos-data collection. They've got to stay around to be able to continually assist, both the investigators and public release files. And then cost! People with these skill sets aren't going to come cheap, so it's a long-term investment. I don't have the answers for any of these at the moment. I like Phil's suggestion of looking to see if there are sub-contractors who are starting to take on these specific talents that we can share as a resource.

Unidentified Speaker: I just wanted to comment that minimizing the risk of deductive disclosure should really be the responsibility of the P.I.'s in the sense that they're ones who understand the survey and the design the best, and they're the ones that can think through how deductive disclosure might work, what's sensitive, and how that can come about. That would just be something that I would point out: that we would never leave that to our honest broker, because he has no idea. That's part of the system. He isn't a researcher, he isn't the P.I., he's not a subject. He's independent of that process.

Unidentified Speaker: Do you tell in your consent form, or when you're telling people there's going to be an honest broker?

Waite: Not so far, we don't.

Unidentified Speaker: So isn't that an issue that there's a third party who's managing this individualized data?

Unidentified Speaker: It would take a half hour.

Unidentified Speaker: I'm just saying, because they are trusting researchers to manage this data.

Waite: What we say is that we won't ever identify – well, never use their name – in conjunction with any of the information they've told us. But having an honest broker isn't inconsistent with that.

Unidentified Speaker: No, I mean, because they're a member of the research team.

Unidentified Speaker: It would be like telling them there're going to be three statisticians, and a-

Waite: One point that I think is really important is that we should not be worried about the risk of deductive disclosure by members of the research team. We have to presume that the people who are scientists working on this are ethical and that they're not looking for the names of particular people to sell to attorneys.

Unidentified Speaker: So why have an honest broker?

Waite: Precisely my question.

Unidentified Speaker: There're two issues. One is being able to identify any person with the data, being able to identify. The other issue is the subpoena issue, outside people being able. That's really two separate issues here.

Unidentified Speaker: And not only that, but this new HIPAA thing. With the HIPAA language, it says we may have to disclose this information to the funding agency or to the IRB. Our IRB was out of my dataset plans last year.

Unidentified Speaker: Well, the honest broker does not protect you against deductive disclosure. I mean, I think we're confusing those two things.

Unidentified Speaker: Exactly. So why have them?

Unidentified Speaker: Do you think you should continue to have one in Add Health?

Unidentified Speaker: Yeah, because it acts as an intermediator between the subject and the researchers. I mean, it's true that a researcher would never purposefully reveal the identity of anyone, but you can easily have very small cell sizes, and we have, for example, one respondent who lives in one state. We don't release the state identifiers either, because you can easily figure out who that person is. Sometimes researchers can unknowingly provide a table with cell sizes that are one person or less than five people.

Unidentified Speaker: Why can't our T.I. be that?

Waite: And how does an honest broker help with that?

Unidentified Speaker: It doesn't, that's what I'm saying. It's a separate issue.

Unidentified Speaker: For me, it doesn't [audio unclear] just linking the specific identifiers, the name, the date of birth, with the data. There's also deductive disclosure, which is taking the data, and saying, 'Okay, there's one black person living in a small town in Nebraska,' or something like that, and saying, 'If I really wanted to, I could go there and figure out who that person is.' A hypothetical risk, but-

Unidentified Speaker: With deductive disclosure there are two issues. One is access, and how easily you can get to the data. The other is the risk of somebody having the motivation to find somebody, and that risk is really small. I mean, I'm the most liberal person on the Add Health team, saying that everybody else is being really paranoid. That's one thing that you have to recognize about me talking about this. Not everybody agrees that that risk is so low, but that's a tiny, tiny risk. We had many scenarios where a father knew that his daughter was being interviewed, and he was waiting outside behind a bush, and he was going to hit the interviewer over the head when they walked out and grab the

computer, because the father wanted to find out things, because he didn't like the boyfriend that the daughter had. And we ask a lot of information. That daughter nominated her boyfriend and told a lot of information about that boyfriend. It could have been the boyfriend's father or mother or whatever. I mean, this is a very tiny risk, but it's a risk that we wanted to protect against.

Unidentified Speaker: And so that's why we mandated the stripping out of the identifiers immediately. They go right to the honest broker. That was really the protection that we were interested in then, but with data release, that's a different issue. I mean, I think the risk is very small, but as [audio unclear] always says, all it takes is one, and you're on the top fold of the New York Times.

Unidentified Speaker: Could I just say that the largest risk is the subject disclosing that they were in the study? That's the risk that you need to caution them about, and that's the best protection, to not disclose that they were in the study. That's huge.

Unidentified Speaker: The risk of someone breaking into our office and finding the genes that we've filed.

Unidentified Speaker: I mean, at some point, however far down, however many hoops you jump through, there's got to be a trust of an ethical researcher, and adding additional groups? I don't think it's really the solution.

Unidentified Speaker: Well, we set up contracts, and exactly, you have to trust in the contract.

Unidentified Speaker: But we are asking for permission to store blood for potential future use.

Unidentified Speaker: Well, that's the thing. If you ever want to do an analysis of that blood, would you have to go back and get additional consent from people?

Unidentified Speaker: Well, we're debating with that with the IRB.

Unidentified Speaker: I think that the IRB is requiring that now. In which case, why even bother getting the DNA?

The Big Picture: What is biomarker collection good for?

Speaker: John Lantos

Kipling has a short story called "The Eye of Allah." It's set in the 13th-century England. An English monk travels to Spain, which has recently been liberated from the Muslims, and he comes back with this amazing device: the Eye of Allah. If you hold the Eye of Allah up to a drop of water, people can observe horrible impish shapes in the drop of water, and a scientist/philosopher who's at the abbey where this monk brings it back, says 'this is fabulous!' It will bring truth to the world. But the abbot looks at this thing and states categorically that this is not about truth, it is a form of magic. He takes a hammer and he smashes the Eye of Allah, and says, 'it would enlighten the world before its time.'

What does this have to do with anything? The discussions here today make me think that the ethical issues surrounding honest brokers are the ethical issues of forbidden knowledge. We want to gather information about people, but feel that somehow neither we nor others should have ready access to the information. We think the information is going to bring good to the world, but fear it might bring harm. So we sequester it somewhere where we can filter out the good without subjecting ourselves to the harm. It seems like the challenge is figuring out what those goods are, and what those harms are.

The honest broker concept is sort of a first stab or a shorthand at imagining this filtering process. DSMBs [data and safety monitoring boards] play a similar role. In the case of DSMBs, researchers agree to deny themselves access to data but they ensure that somebody should see the data.

The term honest broker, though, seems to come from the business world. Imagine that there are two parties in conflict, seeking mediation. An honest broker is kind of a mediator or a deal-maker, somebody who's supposed to be disinterested but has integrity, and is thus able to make tough decisions. From what has been described here today, the honest broker in a research context is less a mediator between researcher and subject, and more a mediator between this researcher/subject dyad (that we research-types all think is one without conflict) and these outside parties who are somehow going to interfere with our benevolent activities on behalf of our subject/clients – who, in our view, we are only trying to help.

Interestingly, the rest of the world doesn't see us that way. The rest of the world is highly suspicious of researchers, who are seen as having our own agendas. All the grousing about IRBs and lawyers, I think, is really grousing about this societal perception of us, that they don't see us as the benevolent people who we see ourselves as. So in some ways, the question is whether the honest broker is simply a way to get around sensible and morally defensible safeguards set up by the current system of research regulation by putting the data offshore where they're not going to be able to get at it.

Alternatively, it could be that the system of regulation itself is so deeply flawed that we're forced to jump through even more hoops to get to the moral good that we see, but that these bureaucrats at OHRP or the people who passed HIPAA just don't understand. By this view, we researchers are the forces of light and the regulators are trying to stop progress.

I'm not sure what is the answer, and I think that to assume that we are so good could be a little dangerous. Let me offer two paradigm cases of research, one that I think is good and that wouldn't be allowed today, and one that was bad and was allowed. Both have to do with sort of biomarker-type

things. One involved a biomarker, a syphilis test on a bunch of black men in Tuskegee, and the researchers wanted to study the natural history of the disease. They collected a lot of demographic information, but they didn't reveal all that information or the interventions that were available to the subjects, because it would have interfered with their long-term research goals, which were to describe the natural history of syphilis. This has become the paradigm case for research abuse. And, to the extent that some of these studies are going to be identifying predictors of Alzheimer's, predictors of cancer, predictors of early death, predictors of hip fractures, etc. To the extent that researchers are not going to do anything about these diseases but, instead, are just going to collect the data, put them offshore in a data bank, and watch our study subjects until they've got a lot of hip fractures, the study looks disturbingly like Tuskegee. I think there's a big moral problem here, and I'm not sure what the solution is. But to say, 'All these safeguards about access to data are simply bureaucratic hurdles to get past' overlooks that.

On the other extreme, imagine that there was no such thing as a birth certificate and a researcher today proposed developing one as a useful public health tool. I think most IRBs would reject such proposals! After all, they would involve massive HIPAA violations by collecting the name of the name of the mother, the name of the father, the diseases, the birth weight, all de-identified and put into a publicly accessible repository with no safeguards on anything that any researcher who wants to could get access to. It would never get through today! And yet, it seems like it's a useful source of information that helps us understand lots of different things about infant mortality and other things. In some ways, birth certificates were thought to be acceptable because they came out of a different research paradigm. They weren't thought of as clinical research or biomarker research. They were thought of as social welfare research.

I'm going to conclude by thinking about different models that might apply to this whole biomarker field, and understand maybe why some of the conflicts have to do with which research paradigm you try to put it into. The two obvious ones, it seems, just from the talks and the people here, are the clinical research versus sociological or demographic research. Clinical research is usually done by doctors. It involves hypothesis-testing, and the hope is that it will lead to some improvement in treatment, some intervention that's going to benefit the people in the study, or people just like them in some sort of pre-conceived way. For example, antibiotic A is going to be better than antibiotic B, or cancer treatment A better than cancer treatment B, or we're going to identify risk factors that we can then intervene upon to correct.

Social science research, by contrast, is more about trying to describe the world. To do it, we collect large amount of data, but without a specific hypothesis to test. We then analyze the data try to see what truths we can tease out of it.

When you get into the clinical model, you get into all the assumptions about what doctors owe patients as moral obligations, which I think are very different from what sociologists owe research subjects. A lot of the problems in biomarker research come from the fact that it is not the doctors doing the research, but the researchers are doing things that doctors usually do, like drawing blood, taking blood pressure, doing physical examinations. It is problematic to assume that people are going to understand that the researchers are not really doctors and that, instead, they are more like those people who call you up on the phone at dinnertime and just want to ask you questions.

There is a third model that is sort of in-between clinical research and social science research. It comes more from anthropology than from sociology: the participant-observer. Someone who goes into a setting, a culture, a sub-culture, to figure out what makes the culture work, and sometimes find themselves in a position of observing things that are disturbing, troubling, harmful to people, and about which they might be able to do something, intervene. They have a conflict: do you screw up your study

by intervening and therefore altering the very things you're trying to find out: what is this culture? How do these people deal with these bad outcomes? Or do you just wait and watch and live with the guilt that something bad happened that you might have been able to do something about. There's a lot written about that, too.

So, in conclusion, it seems that the task is going to be to sort of articulate a new paradigm for this field that really bridges the two in a way that no other kind of research that I can find quite has done before. The more you get into sophisticated biomarkers that involve bodily invasion and other medical-type interventions, the more complicated it's going to become. It seems like the IRB here is the Greek chorus. They are looking at all this going on, and they are saying, 'Whoa! This is disturbing. This is troubling!' They don't know what to do about it, and each one responds differently.

It seems like a practical response to this would be to convene some sort of study group to look specifically at the ethical, legal, and regulatory issues, a group that involved lawyers who were knowledgeable about this, but also involved people from the federal government, OHRP, and the FDA, to talk about the regulatory issues and come up with some sort of statements, guidelines, principles that researchers in this field could have that had sort of prior approval, or at least prior review, or at least prior criteria. Something that had been reviewed and developed by all these people who, it seems to me, are really going to have to work together to solve this one, and setting this up as an us against them sort of thing is just not going to work.

McDade: I think one of the things that you should keep in mind is that, with a lot of the biology that you're talking about here, some of the minimally invasive methods we're interested in applying, the biology is actually a red herring. The most sensitive information is not biological. The exceptions include HIV, some very sensitive issues around drug testing, or whatever. CRP levels are not sensitive information. Cortisol, not sensitive information. Who's sleeping with whom is very sensitive information. So the issues here are not unique to biology, and I think sometimes we need to remind ourselves and our IRBs that we're not adding an additional layer of complexity here, or an additional layer of sensitivity. That's going to vary on a case-by-case basis, but that's something to keep in mind.

With respect to your opening comments, with respect to the fact that we're collecting biological information and letting people go and seeing what happens, that is a potentially very paralyzing stance I think that we need to be careful to avoid. I'm reminded by a series of studies that were conducted in Latin America and Guatemala by [AUDIO UNCLEAR] and colleagues in the seventies, where they went into a village, and they supplemented a group of pregnant mothers and infants with protein. Everyone knew that kids need food, and food is a good thing. Even at that time, that study was on some ethically ambiguous ground, because they gave food to some villages and not others. You would not believe the public health impact that study has had today, because they have documented across two generations the impact of food early in life. They documented that giving foods to kids while mothers were pregnant and kids were in their first two, three years of life had a dramatic impact on their own reproduction, their cognitive potential, and their health throughout their lifespan.

So we look back, and we could never do that study today, but it has really informed our understanding on a global level of what major public health issues are about. Of course we need to be careful weighing the costs and benefits, that's what careful evaluation and research is always about, but I don't think anything has changed. That's what I'm saying.

Unidentified Speaker: I think it's really important to keep in mind the two models, the clinical model and the social science model, but I think that we can overdraw the comparison. For example, in the case study that you presented, did we really need those biomarkers to know that something might be going on there? There were bruises, the guy smelled of alcohol, the woman was unkempt. A survey researcher going in there with just a supposedly innocuous questionnaire, wouldn't that person have had the same kinds of issues? So, again, I think it's important to recognize the different perspectives that we bring, but I'm not sure, like Thom, that it's all about bringing more data in, or a different kind of data.

Unidentified Speaker: I guess what I'm thinking about more is the kind of stuff people were talking about yesterday afternoon. I mean, if you can do a Sniffin' Stick screen that predicts better than any other test the likelihood that somebody's going to develop Alzheimer's, and you're not going to tell them, that's a different sort of problem, I think, than either the STD or who you're sleeping with information, or a questionnaire.

Waite: But I think you think we're better than we are at predicting. So, 'predicts better than anything,' might mean that if somebody sniffs the fish, that increases the chances they're going to get Alzheimer's, or your predictive power, from 5% -- you predict 5% of the cases right, to you predict 6% of the cases right. You wouldn't tell anybody on the basis of that. It would be unethical to do that.

Lantos: So, to the extent that your studies are useless, they're ethically untroubling.

Waite: No, they're useful on a population basis, but they're not useful on an individual basis.

Lantos: But that's true for a lot of clinical tests now, as well.

Waite: Right.

Lantos: I mean, what you're doing is telling people that their odds of having something have gone up, and the question is when you've crossed a threshold where withholding that information becomes ethically problematic.

Waite: And basically, we never do, unless they're on their deathbed now. We never do. The other thing, and this is just sort of the history of demography, you said birth certificates were a research project – they were never a research project. They're vital statistics, it's the way for the government to keep track of their citizens. You know, who flows in, who flows out, it's just a flow analysis. And all this other stuff was just extra.