CCBAR

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

CCBAR, NIA, and Centers on Aging Biomarker Network Biodemography Focus Session

Held adjacent to the 2008 Association for Psychological Science Meeting

Editors: Andreea Mihai, Ada Gomero, Emily Rose Wenrich, Dr. Natalia Gavrilova, and Dr. Stacy Tessler Lindau

April 17, 2008

National Institute on Aging

center on demography and economics of aging

THE UNIVERSITY OF CHICAGO
MacLean Center for Clinical Medical Ethics

Population Research Center at NORC and the University of Chicago

Table of Contents

Emma Adam: Northwestern University

Natalia Gavrilova: University of Chicago

Noreen Goldman: Princeton University

Kate Cagney: University of Chicago

Arun Karlamangla: University of California—Los Angeles

Diane Lauderdale: University of Chicago

Stacy Lindau: University of Chicago

Jack McArdle: University of Southern California

Thom McDade: Northwestern University

Carlos Mendez de Leon: Rush University Medical Center

Teresa Seeman: University of California—Los Angeles

Richard Suzman: National Institute of Aging

Phil Schumm: University of Chicago

Rob Stawski: Penn State University

Rosie Sud: National Institute of Aging

Robert Wilson: Rush University Medical Center

Arun Karlamangla: Socioeconomic Status & Bone Density

Seeman: Okay, well, the first session is going to be a sequence of presentations by Arun Karlamangla, who is a colleague of mine at UCLA, and Rob Stawski. And both of them are going to be presenting data that we've recently started analyzing from the latest MIDUS data collection. And, Arun is going to start off talking about some strategies he's been developing for looking at more cumulative life history kind of exposures and some new ways of looking out comes, particularly looking at the bone. Arun?

Karlamangla: Thank you very much.

Lindau: And I'll mention again for those of you who just came in…we're recording the session so that we'll have the proceedings, but before we make them more widely available, you'll have a chance to review and edit your parts.

Karlamangla: So, this is research in progress talk. It's not really research in progress. It's very preliminary. It's preliminary results from preliminary analysis of incomplete data.

Karlamangla: So this is preliminary analysis that was part of a grant that was resubmitted in February – an R01 proposal to look at psychosocial life history and bone health using three different data sets, which today, I'll just talk about some analysis we did on the MIDUS data. So, MIDUS is mid life in the United States. This is one of the studies that Richard was talking about that Burton Carol had been heading. MIDUS was initially started about ten years ago. Right now, we're doing the second wave. The second wave is almost complete. A couple of these studies – the biomarker study in particular – the daily diaries, [inaudible] are still collecting data. So, this research in progress talk represents our preliminary data analysis using the first 500 participants in the biomarker study and with 1200 or so participants from the daily diary study. So, with respect to bone, why are we interested in the life history approach? Because osteoporosis, though it's a disease of older individuals, it's really a disease of fragile bone. But how strong your bone is depends on how strong your bone gets when you're a young and middle aged adult and how it declines - how this trend declines when you get older. Since how strong your bone gets in middle age might depend on your life circumstances during that time, it made sense to look at not just psychosocial circumstances and old age, but also that when you're growing up and in middle age. Now, fortunately, MIDUS had data on psychosocial factors in childhood and ten years ago, so we wanted to combine all of that information to look at bone health in later age. The hypothesis was that the level of peak bone strength achieved in young adults is affected by psychosocial factors. The maintenance of bone strength in middle adulthood is affected by psychosocial environment. And, of course, when you're older, your psychosocial environment then affects the rate of decline.

Suzman: Can I ask a question in my absence?

Karlamangla: Yes.

Suzman: Did you study Prilosec/Nexium consumption and how it affects vitamin D? Calcium absorption?

Karlamangla: Yes, yes. We haven't looked at the data, but that information was collected, yes.

Suzman: Brilliant.

Karlamangla: So, the point he was making was that we asked older women, especially to take calcium supplements. The most commonly used type of calcium supplement, calcium carbonate, its absorption is dependent upon having acidity in the stomach. And, if you take medicines like Nexium and Prilosec, which reduce the acidity in the stomach, the calcium does not get absorbed. So, he's talking about vitamin D as much as calcium. All right.

And then, in this proposal, we're also going to look at biological pathways, and looking at different biological systems that have been implicated in bone health, including the hypothalamic and pituitary-adrenal axis, the sympathetic system, and inflammation. And data regarding these biological systems are being collected in MIDUS in the second wave in the biomarker study. The other thing that is unique about this approach was that we are certainly going to focus not just on bone density, but on bone strength, and assess that more comprehensively using bone density combined to bone size in relation to body size. Bone density is the usual marker of osteoporosis. But, fragility in the bone depends on more than just density. Now, thinking as an engineer for a moment, the structure or strength of any engineering structure depends not only on the material properties of the material that goes into the structure, but also the size of the structure. Now, we haven't been using size in assessing bone strength, but there have been several research studies that have looked at the impact of bone size. Now, there are also these bone turnover markers that look at the levels of bone resorption and bone remodeling. So, we can use that to assess the current metabolic balance in bone whether bone strength is increasing or decreasing. And, of course, fracture instance.

So, coming to the preliminary analysis, so this is what I'll be reporting on. The preliminary analyses of SES life history taxonomy, and topology, really, cortisol-diurnal rhythm, and bone metabolic balance. So, this is how we created the life history taxonomy. We have these different measures of bone metabolism. So, these, are the different SES measurements we have. We had some information on childhood socioeconomic status, in particular, parental education. There are more variables we collected. This is what we chose to use for these analyses. There is your own education, and then your income nine years ago in the first wave of MIDUS, and your income now. So, we've combined all of that to create life profiles. Now, at the bottom are the people who have consistently low SES, which means they were low – consistently low – in each of those four categories, and then the obvious "best" profile – the people who were consistently high, above the median in all four variables. And we can see that about a quarter of the people made that group.

Now, how you rank all of the groups in between depends on your hypothesis – whether you think your current level of SES is more important than the past. Now, one could do exploratory work and figure that out, but in the absence of that, we decided to go this way. So, we have "high income now, but not always," "low income now, but not always low as the fourth group," and you'll see that there was an order in terms of associations with outcomes. Look at cortisol.

This is from the daily diary study. Participants were asked to measure salivary cortisol at four different times in the day. Once when waking up, 30 minutes after waking to try to capture the initial early morning rise, then at lunch time to catch some of the initial decline, and then at bedtime levels: C1, C2, C3, and C4. We are still working with Dave's group and Rob here very actively, and we're trying to figure out how to best characterize these diurnal rhythms and capture information that would be useful to us. Rob's going to talk more about that. For these analysis, we looked at three different things. Oh, by the way, these cortisol [inaudible]. The next thing we looked at was the pre-awakening level, so when you wake up, what is your cortisol level compared with [inaudible], that's C1 minus the C2. And the awaking response is the 30 minutes after waking minus the waking, so how much do you go up after you wake up.

Woman: What does "recovery" mean?

Karlamangla: Recovery is when you come down from that peak to the bedtime level. Whether you come down quickly or come down slowly.

Okay, on the bone side, we have these, apart from bone density, we also have these turnover markers. There are two kinds of turnover markers. One that reflects the level of resorption. This is bone getting absorbed back. That's NTLT. NTLT stands for "NT-local time" of procollagen. This is a precursor molecule for collagen. And there are two markers of bone formation. Bone-specific [inaudible], phosphatase, and a C terminal polypeptide. Now, it turns out, both types of markers – the resorption markers and the formation markers both are elevated with this bone turnover. So, when you're building bone, it's not just a question of putting on bone. You absorb from the inside of the bone. You think of the bone as a cylinder. You absorb from the inside and add onto the outside. So, both are constantly active – there is absorption/resorption going on at one time. So, looking at any one kind of marker is not adequate because it doesn't tell if there is net growth or if there is net decline. So, we have to somehow create a bone balance index. This has not been done in the literature. People look at one or the other because they assume that postmenopausal women are generally declining, so it's enough to look at the level of one of these markers, if the marker is high, it means that there is more decline. So, that's what people have been doing. But, when you are comparing people across age groups, it's not a smart thing to do. So, we try to create a bone balance index to see whether formation was greater than resorption or to get at an accurate level.

So, to do that, this is what we did. We created T scores for absorption…formation markers and absorption markers. A "T" score is with respect to a normative group. So, men under 50 years of age begin to decline, and women who are not going through menopause yet – pre-menopausal women – were the two normative groups. Gender-specific normative groups, so we find a mean and a standard deviation in this population within these groups and used that to create T scores with everybody. So, a T score is like a Z score, except you don't use the mean and standard deviation for your group. You use the mean and standard deviation for a normative group. Okay, then, if you look at the average T score with the two formation markers, and subtract the T score from the resorption markers. You have a balance index. So, we don't…since we have not captured absorption completely, or resorption completely, one cannot say that this balance is constant or growing, or that this balance index is negatively declining. All you can say is within a population, one can compare people with balance index, if

you want higher balance index, it means more favorable towards absorption compared to distress.

Lindau: Are you receptive to questions during your talk, or would you rather have questions at the end?

Karlamangla: Either way.

Lindau: Teresa?

Seeman: I'd say go ahead.

Lindau: Okay. It had to do with that last slide. So, first of all, there are normative population data on the…for these markers in young men. Is that the case?

Karlamangla: Oh, no, there aren't. So, what we did was we used the MIDUS data to find the mean and the standard deviation. So, at least to my knowledge, there is normative data for bone density, but not for these markers.

Lindau: Because we're struggling with a similar issue with salivary sex hormone data. So, we have sex hormone data on older men and women, and there are no population norms in those age groups. So, this is an interesting method that may be potentially translatable.

Karlamangla: Yes, there is the same problem in osteoporosis. Most studies have looked at older people where the decline occurs, so we don't have normative data. There are a few studies from England where they have some information in younger folks, and most people who have looked at younger folks have looked at clinical trials to see if their bone formation increases if they exercise more. There isn't really normative population-level data that we know of. So, this might be the first time that we collect that.

Lauderdale: But this idea is really familiar in bones because the bone mineral density – this is like an automatic thing that comes out of a machine…

Lindau: …for women, right.

Lauderdale: Well, for men and women there are normative values, and by race…

Karlamangla: For bone density.

Lauderdale: For bone density. Once you get women [inaudible] for bones, this seems like, of course you'd do this.

Lindau: Right, like Z scores, are clinically used scores that help us decide whether to assign treatment or not.

Lauderdale: Right, for bones, Z scores are both related to the same age group and also related to the people in their 20's.

Karlamangla: In fact, that's what I had used to get the T front, so Z and T in bone density – Z refers to these kinds of scores with respect to your own age and gender and ethnicity group. T scores refer to your gender, ethnicity, but at a younger age and original peak bone density.

SES and Bone Density in the Hip

	Correlation Coefficients
Parent Education	0.08
Own Education	0.09
Current Income	0.18

McDade: So there's obviously this balance between formation and resorption, and so ostensibly, if the formation is greater than the latter, you get growth and vice versa. Does the relative efficiency and balance between those two change with age?

Karlamangla: Yes, well, I'm not sure I completely get the question. But if the question is, how these values change with age. That is part of the validation of these indices. I'm not sure I included that because I was getting to the essence of the associations between these indices. But, in our validation peaks, which we presented at the American Geriatrics Society a month ago, we looked at these balances indices to see if they were consistent with what we expected. In a sense, does this balance become more negative in menopausal women compared with premenopausal women? Does this balance get more negative as you get older in men? And we see exactly that. However, if we look at the formation markers themselves or the absorption markers themselves, you don't see that. They seem to go up and down as you go through menopause, and the same thing as men get older.

McDade: So, what you found was that not just that the balance changes over time, but is it relative to the rate of bone decline?

Karlamangla: Ah, so, the construct. So, in other words, do these markers reflect the level of absorption the same way in different age groups? Well, I don't know that.

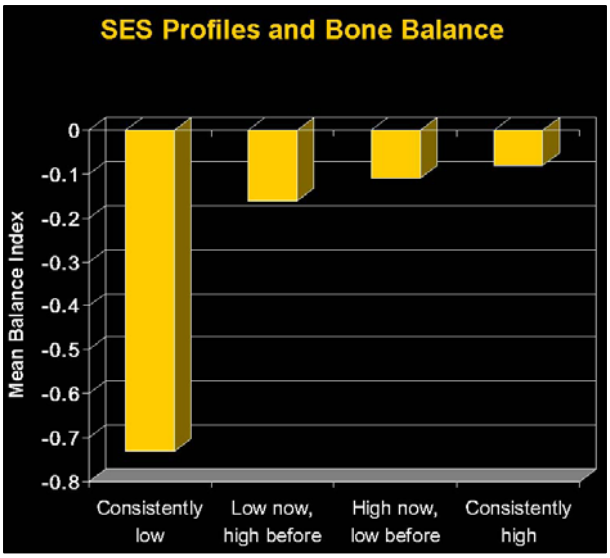
All right, so, the first thing I'll show you is…let's say we do the usual thing, which is look at the current SES and bone density. We get weak to moderate levels of correlation only the correlation with current income was significant. Again, we're still using the [inaudible], but bone density isn't even available for 500 individuals because we've got money to measure bone density a bit later than we got money to measure the…actually, the general markers [inaudible], but bone density is available for about 117 folks. So that was very good, very good. So, this is what happens when I look at the SES profiles of whole groups we talked about…it's consistently low and consistently high and look at the bone balance index, and we see this nice ordering with the people who had the low SES – the people who had the most negative bone balance in the MIDUS 2.

Lindau: So, how do you explain that?

Karlamangla: How do I explain that?

Lindau: Yeah, why is that?

Karlamangla: Well, I guess the usual stress hypothesis of people who had low SES are stressed and then I'll show you that the corisol levels are also dysregulated, so cortisol may be the pathway to more turnover.



Man: May I ask you a question? Wouldn't you attribute that to nutrition?

Karlamangla: It could be nutrition.

Man: But that isn't the first guess about this? I mean, about bone growth…

Karlamangla: Yeah, because most studies that have looked at…I'm not one to say, but other outcomes, there are just things that nutrition doesn't…the SES associations don't go away.

Correlations: SES and Cortisol Rhythm

	Own Education	*Current Income*
Nadir (lowest bedtime value)	-0.04	*-0.10**
Highest pre-waking rise	*-0.06**	-0.01
Highest awakening response	*-0.08**	-0.04

** p < 0.05*

Woman: You could actually look at that because you're consistently low on income but that's not, like, really, really low. So, some of those people might have low income, but it's certainly low enough so that really bad nutrition is probably not an issue, so you could look at that.

Karlamangla: And this was America, not the third world.

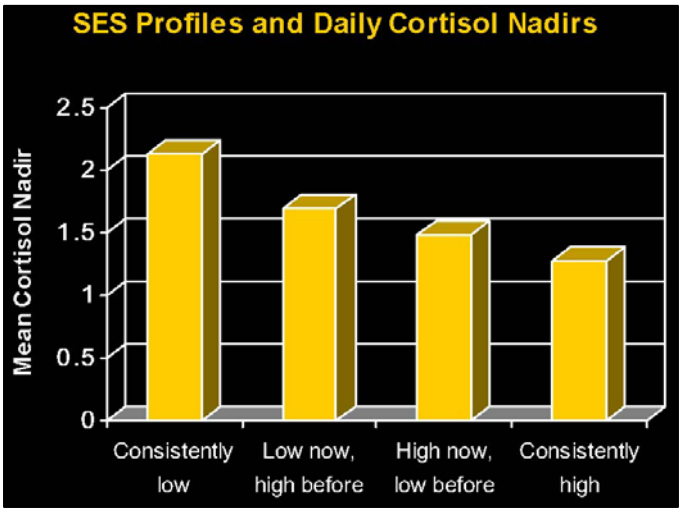
Woman: No, but it could be part of it, but we have other data to suggest that cortisol…

Karlamangla: Okay, and moving on to SES and coritsol. So, if you look at the straight correlations between socioeconomic variables and the coritsol [inaudible] that we talked about. There are these correlations that are weak, at best. And then, if you look at the SES taxonomies and the nadirs – the cortisol nadirs, again, you'll see this ordering with the people that have

consistently high SES and then the lowest nadirs. So, you expect, in a good, well regulated cortisol rhythm to have a nice, strong, robust increase in cortisol levels as you wake up, and then they go back to a low, resting level at bedtime. And, that would be good. And, if your resting level is higher, that's a marker of stress. So, what we're seeing here is that the people with the consistently high socioeconomic status through life have the lowest cortisol nadirs too.

As I said, this is all very preliminary. So, to summarize, SES life histories appear to have larger influences in these bone balance indices and coritsol nadirs than just the current levels of SES. Now, of course, we haven't said anything about strategies we have…we can guess this in different ways. We're not there yet. We're waiting for the pool of data – that will be in the summer, and we are working with Rob Stawski and Dave Almeida on cortisol rhythms.

Woman (83.0): I'm not sure if I missed it. Did you present the association between the cortisol profiles and the bone density measures?
Karlamangla: No, not yet. This is what was happening. So, these are the first 500 people from the biomarker study and the first 1200 from the daily diary study, and there wasn't that much overlap. The people who did the biomarkers are now doing the daily diaries study, and vice versa. So, when the full data set comes, we'll have enough overlap.



Woman: Well, we didn't do it here, and Rob, you're going to talk about the actual slopes. We'll show you. We've been also looking at the slopes - literally measured during the day across the different timepoints, across the…

Karlamangla: This is what a cortisol rhythm looks like, and I'm sure theirs will be exactly like this. There's a nadir somewhere in here, and it starts rising up. This difference here is what I call the "preawakening rise," and it's the same as this slope here.

Woman: Right, exactly. Yeah. Except you're using the lowest point, not the night before across all the days.

Man: Yeah, why did you take the lowest point from the full four days instead of the average of the four days?

Karlamangla: Because we had to truly get to the nadir. Because sometime when you go to bed, you still have a little bit of reaction to stress that you had throughout the day. So, hopefully, one of the four days, you're more relaxed.

Woman: This is trying to give you the benefit of "what's the best you achieved." What was the lowest you managed to get down to over the days we saw you?

Man: And that the average of the four is more characteristic of the person than that one day that they happened to, you know, go down…

Woman: And the low SES may just have four bad days, which, you know...

Karlamangla: That is very true. It's very difficult, and we'll see this when Rob talks about this too. It's very difficult for us to tease apart if the effect we saw was really the effect was here, or there…they're all so closely tied together. When you talk of recovery, you talk about whether they come down quickly or they come down slowly, or but if you just look at the total decline from here to here, that's related largely to how you went up. So it's quite difficult to characterize. Some people have looked at the curve.

McDade: Arun, when you showed your first results slide of the bone values. The two middle groups, the low-high and the high-low. I was expecting that those would have been switched, actually. The low SES earlier in life would have had more impact on the bone parameters because of the developmental story that you're talking about.
Karlamangla: Good question. The reason it did not is because you were thinking of bone strength and I was measuring bone metabolic balance later in life. So, my metabolic balance today may not remember what happened ten years ago. It's more of a factor of my SES today. If you look at bone density, you find them switched.

McDade: And does that suggest, since the cortisol results map nicely onto the same pattern – sort of the same dose response relationship. Does that suggest that cortisol mediated bone resorption processes might be more important later in life, or later in life it's bone deposition.

Karlamangla: We don't have cortisol earlier in life, but cortisols now.

Man: So, if you could look at the relative efficiency of the bone absorption/resorption process, you would have equitable standards.

Karlamangla: Right. Well, bone density today is an integrate of all that happened in the past while bone metabolic balance today is more indicative of what's happening now. So, if you have bone density today, tomorrow, and in the future, that might maybe validate one or the other. But crosssectionally, it is difficult to say whether there will be a strong correlation between…

Man: But if bone density today is more a reflection of lifelong process, wouldn't you then expect that to have better association with your life course history? That is what you'd expect?

Karlamangla: So, we've been more than meaning to look at bone density, bone size, and bone metabolic rate. We've done that. We've validated that in practice. In the proposal that we've put together, we're looking at this life history approach more on bone strength, and then bone metabolic balance's ongoing changes' particular future.

Lauderdale: Yeah, I've done some work about ten years ago that's kind of related to that, which was looking at hip fractures in Medicare. So, there's a geographic variation in the US in hip fractures for men and women. The patterns are a bit different for men and women. Basically, for women, there are really high rates in the southeast, very low rates in the central Midwest and the northeast, and sort of medium in the west. And for men, there are also really high rates in the central south, like Texas, Oklahoma, and a bit in the west. And what I did was I looked at that geographic variation and to what extent for men and women separately it was attributed to where they lived when they fractured their hip, versus where their social security numbers were issued, which is encrypted in the first three numbers of the social security number. For that generation, it was something that people got when they were employed. And what I found was that the geographic variation for men was entirely explained by where the social security numbers were issued, not where they lived currently. Whereas for women, about 2/3 was explained by the social security number issuants, and about 1/3 by where they currently lived. This made sense, on reflection, because it spoke to the fact that the factors around where they were laying around on their peak density were terribly important, and then there's just this relentless slow decline from age thirty on for men. Whereas, for women, there are two periods of very rapid change in bone density. There is the early life laying down of peak bone, but there's also the few years postmenopausal, just after menopause. That was another opportunity for geographic, which may just be SES factors that are correlated, it's unclear, to also play a role. But it speaks to, first of all, that men and women being are quite different, and, for both, there is a correlation with early life factors. And, you're quite right, this is not terrible germane to looking at the metabolic factors later in life, but when you get in with peak bone density and fracture, it's a really complicated story of the relative importance.

Karlamangla: That's very nice. So, they've had these studies, which I'm sure you're aware of, where they've looked at how much of your facture risk depends on your peak bone density achieved and how much depends on individual decline. And, of course, I forget what the breakdown was.

Rob Stawski: Modeling the Diurnal Rhythm of Cortisol

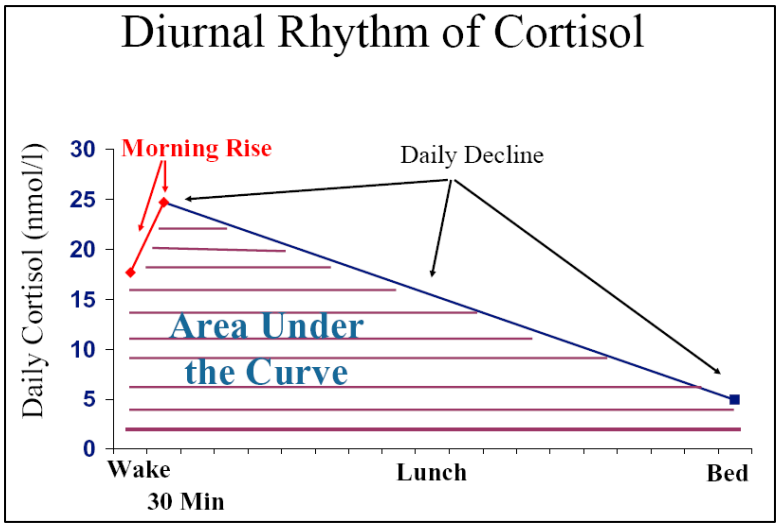
Seeman: Okay, thanks, Arun. Our second presentation from the MIDUS data is Rob Stawski, who is a postdoc in Dave Almeida group, and a superb data statistician methodologist.

Woman: He's not just a stand-in for Dave Almeida.

Stawski: At least hope that we're not going to do a dog-and-pony show for you all today. If I'm kind of standing in the way of anybody being able to see the screen, somebody just sort of cane me off to the side or something like that.

Woman: You can just keep moving from side to side.

Stawski: I'll just kind of stand over to the side. What I want to focus on today is just providing you all with the idea of an assessment protocol for daily salivary cortisol in the national study of daily experiences – what Arun was talking about during his talk. This is a multiple occasion daily sampling of salivary coritsol, trying to capture diurnal rhythm of coritsol in people's natural environments. This is one of the subprojects – this is the daily diary project of the MIDUS 2 study. What I'm going to be presenting on today is our approach to modeling diurnal cortisol rhythms when we have multiple occasions – mutiple cortisol samples per day, but then we also have multiple days of cortisol for a given individual, then we have a plethora of individuals.



So, as Arun alluded to, there is a fairly robust diurnal rhythm to cortisol where we see a fairly precipitous peak or spike in the first thirty minutes in the day and then a progressive decline throughout the afternoon and into the evening, and this was kind of our "working pattern" that we were trying to recapture with our cortisol sampling. Some of the analytical challenges we faced with the picture you just saw are first and foremost that the diurnal cortisol trajectories in and of themselves are not linear. You see an increase and then a decrease, so a nice straight-line function isn't going to fit this data across a given day. Then, we considered the fact that if you're collecting cortisol data on multiple individuals, Teresa's diurnal rhythm is going to look much different from mine, which is going to look much different from Arun's. I think everybody's in here is probably going to look different from everybody else. Then, also, if we get into sampling cortisol across multiple days for individuals, an individual's diurnal cortisol pattern might not look the same from day to day. So, there's not

only variation across the samples within a given day, but also across the days and across individuals.

The data I'm going to presenting on today comes from a nationally represented sample of participants from the daily diary and we have people from MIDUS, which isn't going to be the focus of today. There were 1124 individuals who provided these multiple daily, multiple day cortisol samples. The mean age of the sample was 57, so fairly broad age range: 33-84. Sample was 55% female, and it's a fairly well-educated sample with at least 70% having some college across their life. So, the protocol for capturing salivary cortisol was that individuals were sent basically sixteen salivats, and they were asked to take a sample first thing in the morning from when they woke up, thirty minutes after waking up, just before lunch, and just before bedtime. In here you can see that the sample average times these saliva samples were provided, as far as indices of between person and within person variation. The between person standard deviation reflects individual differences in when these cortisol samples were taken. Within person standard deviation kind of reflects, for a given individual, how much variation there was when they took their wakeup sample across the four days, so we see that there's, in general, maybe a one-hour variation for a given individual in when they took their samples across the four-day period. So there is variation, between and within persons, when they took these particular saliva samples.

Sample Characteristics

- Nationally representative sample of participants from the daily diary and cognition Projects of MIDUS II (N = 1,124)
- Mean Age = 57 (SD = 12, Range = 33 – 84)
- 55% Female, 45% Male
- Education
  - 30%: HS Diploma or less
  - 51%: Some college - 4-yr Degree
  - 19%: More than 4-yr Degree

So, for the data we're looking at today…they have a grand total of 17400+ samples. There are a total of just over 18000 possible samples; however, 546 were miscollections, and there were a few out-of-range values, but this resulted in a total of just about 96% of the samples collected being usable. As Arun alluded to, the daily diary project is still in the field, and we're slowly climbing our way to just around 2000 individuals who completed this cortisol sampling protocol and, right now, we're just around 1700 or 1800, and we have not gotten around to cleaning all of that data yet. But, we'll at least be able to see a first pass at the 1100+ individuals that we do have data from currently.

Salivary Cortisol

- Salivary Cortisol (4x/day for 4 days)

		Mean	BP SD	WP SD
Sample 1	Upon Wakeup	6:40	1.05	0.79
Sample 2	30-Min. Post-Wakeup	7:14	1.06	0.8
Sample 3	Before Lunch	12:40	0.8	1.14
Sample 4	Before Bed	22:25	1.05	0.67

- **17,465** samples
  - 18,288 possible samples (1,143 * 16)
  - 546 missed sample collections
  - 305 out of range values
  - 95.95% usable samples

Lindau: Rob, can you say anything about external validity of the cortisol measures in your study, or how have you gone about evaluating that?

Stawski: Well, this is a good question and one that we struggle mightily with. For a certain subset of the individuals, they received a smartbox, which was basically a box that the salivats were sending in and trapped when people opened and closed the box. I think they were only told to do this when they took a particular sample. People seemed to like the box. You know, maybe it's just something about the clicking that they liked or something, but they open it quite considerably, so, we've had to go back and take a look at a plausible range of times within which these samples were taken. We've been able to recover the same patterns.

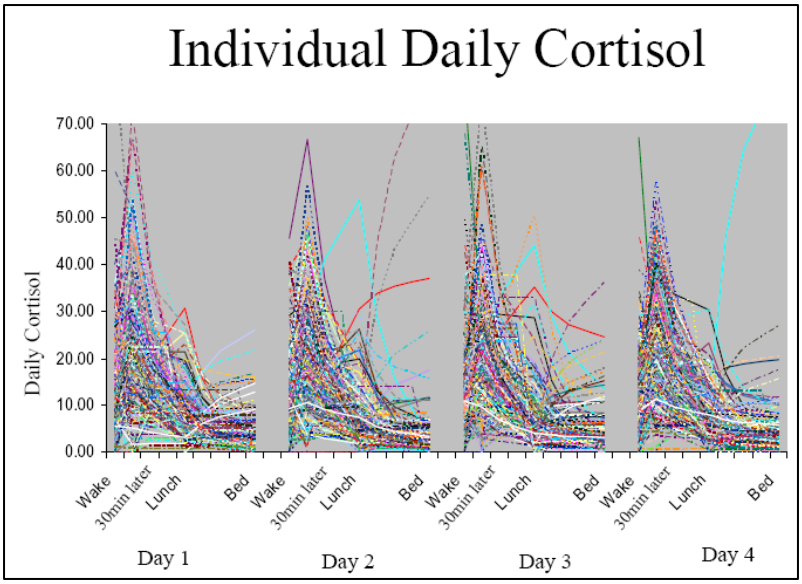
Lindau: What about the data versus other studies that have collected similar cortisol data?

Stawski: Yes, when we have compared our results with other studies and have tried to replicate the studies presented, and similar studies from [inaudible] Kurschbaum's lab as well as some of Art Stone's research. Our slope estimates, at least in terms of ordinary slope estimates, map on quite nicely. But, the magnitude of the slopes is pretty comparable. We have slightly larger amounts of variation in the samples, which is probably largely attributable to the fact that we have over 1000 people, whereas many of these studies have far fewer people. So, in terms of the absolute magnitude of the slopes, we're able to recover what we see from much higher controlled studies and sampling.

Lindau: Are the actual concentrations of cortisol similar?

Stawski: Yes. Not for everyone.

Lindau: Yes, but they are in the same order of magnitude of range…



Stawski: Yes, they are.

Woman: It's going to depend on the lab, though, so, I mean…

Stawski: Yeah, our samples are…

Woman: Clemens, yeah, so, it'll come…

Man: What portions of your 305 and below value observation were "analyzable" low value versus not enough saliva to be able to…vary by age…I'm thinking about the

older folks and the inability to produce saliva. Did you run into that, or did you do something different to…

Stawski: Most of those out-of-range values were actually on the high end, if I recall. So, these were just…abnormally high values. I don't recall how many specifically were at the low end or below detection. In terms of systematic associations with variables such as age and such…with those particular sampling, those 305 values, offhand, I can't speak to that.

Woman: That's a good question. We should look at that.

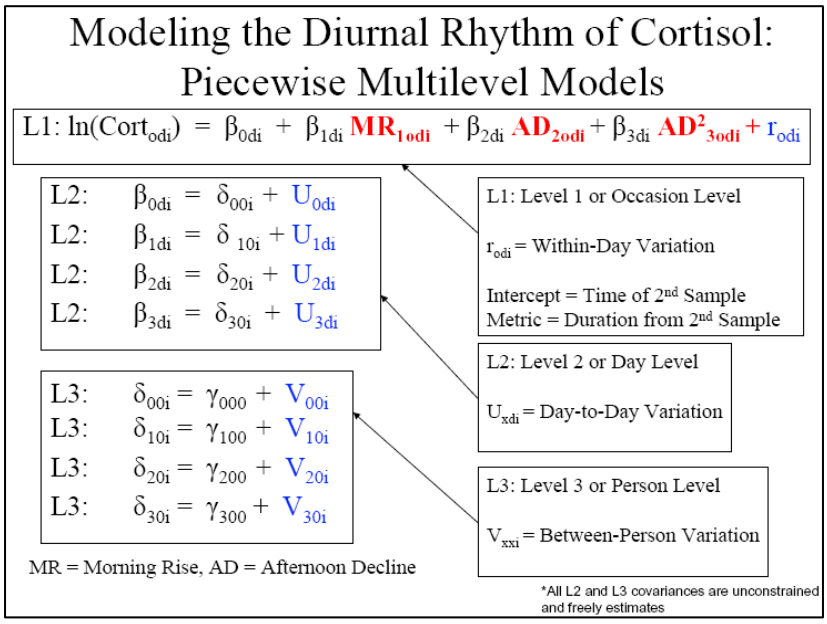
Lindau: You might be interested, the NSHAP data set included salivary specimens from people 57-85, so we had about 1000 people, I think all 3000 may got…maybe 2000, so maybe 800 people in the 75-84-year-old age group, and about 85 or 88% gave us salivary samples that were usable. Those data are publicly available.

Stawski: So, here is just a scatterplot. This is a subsample of about 250 random participants from the cortisol sampling protocol, and you can see that we do see these patterns arise over the course of the day, where there's a spike in the morning and then a decrease throughout the day. But, you can also see considerable individual differences in cortisol levels and slopes, and you also see considerable variation across the day. Individuals are not looking identical from day to day, so this is just a graphical representation of, for better or worse, the mess that I get to work with on daily basis.

So, a little bit about our approach to modeling this diurnal rhythm. We're actually using a piece-wise multi-level modeling, so how we're defining cortisol trajectories is along a timeline. This metric is basically the duration in hours from the second samples. So, we're defining the intercept as the second sample – the people who were supposed to take it thirty minutes as opposed to waiting. And, the morning rise we defined as the slope between the first two samples, so this is the duration between samples one and two. And the daily decline is the slope between the last three samples, so the duration between samples two, three, and four. And, I know that there is some current ongoing faith about how best to characterize the daily decline from the peak to the end of the day, or from waking to the end of the day, and this is definitely something that the interest of the people here presided on and have some open discussion about, but this is how we've approached the modeling of cortisol for our current data.

And, in general, when we're modeling cortisol, and we start modeling cortisol at what we call "level one," which is the level of the occasion. So, this is occasions across the single wave for a single individual. Log-transform cortisol and then we model cortisol as the function of an intercept, so that person's level of cortisol at that second sample. We have slope parameters: the first slope parameter, the morning rise, reflects that duration between the first two samples, then the last two parameters, AB and AB^2 are these afternoon declines. So, we allow for the decline to be both linear and non-linear. There is an initial rate of change, which is the rate of change or the rate of decline thirty minutes post-waking, and then there's some non-linearity. We allow for the fact that cortisol declines throughout the day may not be constant. So, what this captures is basically the general trend in cortisol on a given day for a given individual.

Then, from there, we allow these parameter estimates to vary at level 2, which is the day level. These new parameters here allow there to be day-to-day differences – day to day variation in the cortisol. So, everyone is allowed to have a different value on each day. It's not constrained to be the exact same value across the days.



Similarly, at level 3, these new parameters allow for individual differences in the diurnal rhythm of cortisol. So, it reflects that some people are going to have higher levels of cortisol, as well as steeper or shallower slopes than others. Back to level 2, on some days some people are going to have higher or lower levels of cortisol than they usually do, and their slope parameters may be steeper or shallower than usual.

And, I do want to mention, just briefly, that in terms of fitting these variance components above level 2 and level 3, we do not constrain the covariances. We allow these to be freely estimated. So, all the variances and covariances are unconstrained.

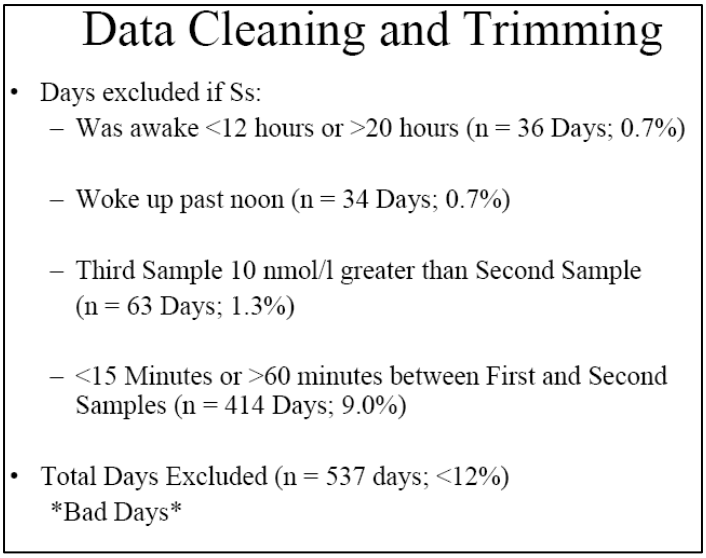
So, I'm just going to go through this slide briefly, but the main take-home here is that the fixed effects retrieve the sample average parameters for cortisol. A bunch of these are in natural log units, so we see the level of cortisol at approximately the second sample of thirty minutes post-waiting, and we see a fairly precipitous morning rise. That's negative because we're working backwards from the second sample. Then we see both significant linear and quadratic daily decline parameters. So, we see that there's a fairly precipitous decline from that second sample in terms of cortisol levels; however, that rate of decline is abate or shallow in all over the course of the day.

Parameters of the Diurnal Rhythm of Cortisol: Means and Variability

	Fixed Effects	Variance Components (Level 3: Person)	Variance Components (Level 2: Day)	
	Estimate (SE)	Estimate (SE)	Estimate (SE)	ICC
Intercept (30 Minutes Post Waking)	3.013 (.014)	.150 (.009)	.039 (.004)	0.444
Morning Rise	-.529 (.021)	.210 (.021)	-	1.000
Daily Decline (Linear)	-.233 (.004)	.008 (.001)	.003 (.001)	0.727
Daily Decline (Quadratic)	.007 (.000)	.000 (.000)	.000 (.000)	0.657

*All parameters significant p<.0001

Intraclass Correlation (ICC): Proportion of variability attributable to reliable individual differences



In terms of the variance components, all the parameters up here are significant. So, what we see is that there's evidence that significant intra-individual differences in both level and the parameters of the cortisol slopes across the day. So, there are individual differences in how steep the morning rise is, as well as the rate of decline throughout the afternoon and into the evening. But, we also see evidence of significant variation across days. So, for an individual, we see variation in their level and the rate of decline throughout the day. There is no significant variation in the morning rise, and we can talk about different reasons that might be, but with only two points defining the morning rise, it's going to be a little more tricky to isolate out significant day to day variation. In the end here, there's a fairly high degree of variability here in the interclass correlation, which is the proportion of variability attributable to reliable individual differences per group or person level, or proportion of variability here. The proportion of variability is attributable to individual differences, so about 66%, 2/3 of the variation can be slope-parameters is reflective of stable individual differences. So, what this suggests is that there are considerable individual differences in diurnal rhythms in cortisol. However, there also is evidence of a fair amount of day-to-day variation for a given individual in terms of their cortisol.

Karlamangla: Quick clarification, so the morning rise is the difference between the peak and the starting level? C2-C1?

Stawski: Yes.

Karlamangla: And daily decline, there are slopes, right? There is an option of time.

Stawski: Correct.

Woman: The morning rise is a slope too, though. It's a rate of increase.

Karlamangla: They are all rates.

Stawski: Yes, yes. These are all rates.

One of the things we've been working on, actually, very closely with Teresa and Arun and all our colleagues at UCLA is what are some issues that are going to arise in this sort of field sampling data as opposed to lab data, where we can actually have better control over the saliva sampling. We've come up with a number of "flags," as we call them, to what might qualify a

day that's invalid, or a "bad day." It's not going to be representative of a true, healthy, normative cortisol rhythm. And, the flags we've come up with here are a day where the person was awake for less than 12 hours, or greater than 20 hours on a given day. It's happened on a very small portion of days, 36 days, which was a grand total of 0.7% of all possible study days. We also saw a small portion of days where individuals reported waking up after noon, and we have also had, then, some days where the third sample, the lunchtime sample, was actually ten nanomoles greater than the second sample. So, instead of seeing a peak and then a decline through the day, these people just kept going up. And, we've been talking about this, whether it's an issue of compliance: people go the two samples switched, in some cases, people are eating lunch before they take the sample, which is affecting their coritsol levels, and then finally, we qualify that there may be some issues with detecting the true morning rise that people take the samples within fifteen minutes of each other or let more than an hour lapse between the samples. And, by far, this particular flag encounters the highest number of days – about 9% of study days are characterized by potentially problematic cortisol sampling between the first two times. People just did not take the samples when they were supposed to.

And, altogether, this leads to a total of 537 days, just under 12% that we deemed as maybe "bad" or "problematic" days. The reason this is important is displayed on this figure. So, what we see is, we take and we model the "good days" versus the "bad days," the ones where a flag was identified. We see that individuals…I'm sorry, the slopes of what were levels of cortisol on these bad days. We also see that the morning rise is considerably less stark. And we also see that the daily decline is considerably flatter on these days. So, on these days where people are deemed to be non-compliant, or there are some problems with the sampling protocol, we do see a fundamental shift in what we're capturing. So, this could just be noise, but we're trying to investigate things that may systematically account for why there were problems with the same kind of stretch and what might explain these sorts of differences.

And, to provide some evidence of validity to this method, we wanted to show that this cortisol sampling can be linked to something theoretically informative. In this case, stress. So, here, what's displayed here are within person predictors of diurnal cortisol. And, in this case, we're looking at stressors that occur within and without family members. This black line here kind of shows the average cortisol rhythm on a non-stressor day. Think of these three lines as one individual. This is an individual's diurnal rhythm on a non-stressor day. The red line reflects what happens to that person's rhythm on a stressor day that does not involve a family member. We see a shift in the decline. They're not as steep on that day. Then, if it's a day where they encounter a stressor with a family member, they look really bad. We see that their cortisol levels are lower and their decline really flattens on. So, basically, it goes low in flatline days where individuals report experiencing a stressor with a family member.

Lindau: Is this excluding the "bad day" people?

Stawski: Yes, yes. This excludes the "bad day" people.

Lindau: But, actually, though, I misunderstood, because I thought where you were going was

that you were trying to make something of those invalid days to see whether it was stress that interfered with the collection of the sample. So, did I misunderstand something?

Stawski: No, I was just putting that up there to show that even though we have a certain sampling protocol that when we can identify deviations from that sampling protocol, it manifests in a shift in what we're capturing and what we're able to display on a certain day. So, that's just kind of showing that this is a concern to be aware of working with this kind of data.

Lindau: But you didn't go on to see whether those people who gave you the invalid days were more commonly self-reporting stress?

Stawski: We have looked at that. It's not just people, but it's days. So, these days are coming from multiple different people, and we're not finding that something like stress is predictive of that non-compliance on a given day. We've sort of kitchen-sinked this with age, education, gender, trying to find out what sorts of things are explaining this.

Woman: Is it more likely to be on the first day that tends it to show up? No?

Stawski: No, no. Not necessarily. No.

Woman: And some of those days include fewer people. So, there are some people who just basically didn't do the right thing any day.

Woman: Right, right.

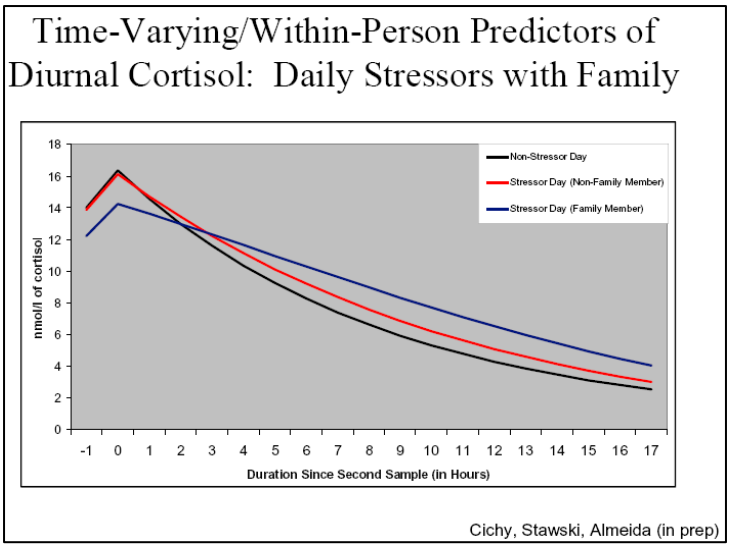
Woman: And so they're accounting for multiple days of the same person…

Lindau: We tend not to go down the path of analyzing that data and publishing those findings, but they're really useful as people design their methods for going into the field with these…

Woman: No, I mean, we found, when we did this, and then you go back and you analyze looking for variables you think ought to predict rhythm. It's a much tighter look because what you're essentially getting rid of is error of when you're saying "this is their morning rise," and they did that second sample an hour later. That's not their morning rise. You missed their morning rise.

Lindau: Well, getting rid of error is really important. But, actually finding correlates that operationally affect people's adherence to the protocol is an important piece of what we talk about at these meetings, but people really aren't spending time writing those papers and publishing them.

Stawski: Actually, just to qualify something about the issue of the invalid or noncompliant days -only three people had this sort of invalid data across all four days. The lion's share of these invalid days were spread out - single days for multiple individuals, so by and large, this was just an individual having one bad day.



Man: And did anything vary significantly with that?

Stawski: The only thing we have been able to really show are correlations with the morning rise. So, predictors of having had a greater than sixty minutes between your first two samples or less than fifteen, and the one thing I have kind of found with that was general cognitive function. People with lower cognitive function are more likely to have non-compliant days. However, I do want to qualify that the correlation coefficient for that is not very large, and it accounts for less than 0.02% of the variance. So, the reason the correlation is significant is one is that flag has the most noncompliance, and we've also got a sample of 1100 people. We're able to eke out a significant correlation, but it's accounting for less than one half of one percent of the variance. So, it's there, but it's still relatively small.

Woman: Just one other idea with respect to the invalid days, or data points, is with the amount of data you have, you can throw people out and it's not a huge problem. For smaller data sets, another thing to do is statistically control for noncompliance. So, for that awakening response, put a dummy variable in saying "this person is five minutes early, five minutes late," or continuous "how far off are they on this sampling," and pull out at least some of that variance.

Woman: Yeah, so the stress with the non-family member is the thing of beauty here because clearly something is happening during the day which keeps them from adapting. The stress with the family member seems to have been pre-ordained because from the moment they wake up, things are looking different. And, yet this is within person predictors. Because, I mean, you can imagine family stress tends to not be a one-day event. However, this is within person. So, what do you make of that?

Stawski: Well, yes, I would agree. The likelihood of being able to fundamentally shift the entire distribution is greater with a family stressor, especially if you're waking up with them. But this remains, even if you control for what happened on the previous day with people, it's actually trying to control for a carry over effect.

Woman: So, is the causal direction going the other way? You're cranky because your cortisol is…I mean, what? It's very odd.

Karlamangla: Your spouse is cranky because your cortisol is high.

Woman: I've got some data where I looked at kind of day-to-day dynamics, and you…low

cortisol in the morning did predict greater fatigue and kind of, pain for the rest of the day. So, there a reason, which is to activate…

Woman: Maybe there's some underlying factor that's both predicting the cortisol and the problems that you see with the family.

Stawski: This is something we've started thinking about more closely now, especially given these findings, is the direction of the effect. Basically, can, you know, having malfunctioning chronobiology on a given day actually, as you said, preordaine, what might happen…are you more likely to get into fights with people?
Woman (16.0): And I'm going to suggest, which has to do with what I'm talking about, that you slept badly. And that's causing both.

Woman: I mean, because we certainly have people who are in this daily diary study who have also answered the more general questions from MIDUS about just the frequency of that kind of stress. Not these particular days. And it affects these cortisol rhythms also, on top of just what's happening. There may just be some recurrent kind of problem, and that that's affecting both the wakeup as well as then when you have a bad day, in which case that slope looks bad as well.

Schumm: I'm curious, what are you plotting here? Are these really within person differences, or are these sort of averaged over the sample as a whole?

Stawski: Well, no. These reflect through time varying within person associations that we separated out individual differences and stressor exposures from whether or not a stressor is experienced on any given day. But these effects are assumed to apply to the entire sample, so we're assuming that these samples are fixed.

Schumm: In other words, what I'm asking is each individual sample member different in terms of of their four days, which ones they had stress on. Some will have it on all, some will have it on none. And is what you're plotting here basically the sort of average values for the model across all of the nonstressed days, for the one thing, and you know, stressed days for the other. In other words, are you showing a population averaged line for each of these "types" – these three "types" of days?

Stawski: So, this would be showing the expected rhythm for a given individual on a given day. So this isn't the population average. Think of this as the three separate lines that made sense for a person who experienced a non-stressor day, a stressor day that did not involve family members, and a stressor day that did involve family members.

Schumm: Sure, except that your estimates of each will come from slightly different subsets of individuals. I guess my point is, if I understand correctly, the true within person differences across these three types of days may even be greater than what's actually seen in this plot.

Woman: Were those obtained from the day-level coefficients?

Stawski: Yes.

Woman: Yeah, so it is a within a person difference that's just laid. That's the average within-person shift from a good day to a bad day added to the population intercept with everything taken out.

Karlamangla: Yes, but the problem is that the people who don't have any stressful days are still in the margin. And the people who have all stressful days are still in the margin, so that's confounding. We did exclude them in calculating differences, so there is simply some vestigial compounding between individual differences.

Woman: It's within-person change that, so the people that had no differences between days wouldn't contribute at all to this model.

Karlamangla: It's not that clean unless you exclude them.

[Voices overlapping]

Stawski: The only way to make it clean, at least in terms of the perspective I think you're getting at, is to identify the individuals who had nonstressor days, stressor days, both kinds.

Karlamangla: And those who didn't have that, leave them out.

Lindau: Did you look at the fourth group that had stress with both family members and non family members?

Stawski: This is one of the things that we have not gotten to yet. We need to rely on some of our coded data that we get from these phone interviews that contribute to the stress data because for the data reflected here, these are coming from the stem questions from the telephone data in the entire interview. So, it's basically, "did you experience a stressor" and "was it with a family member." There's only one specific stem question about interpersonal tensions. So, on a given day, that question can only apply from either a non-family member or a family member. When we delve into some of the coded data, we will get narratives on this. There may be other stressors that occurred that really do kind of lend themselves to being characterized as interpersonal tension and be able to test that, but because we used the stem questions the way we have from the daily diary, we cannot get at that because we do not have separate questions for "did this stressor occur with a family member," "did the stressor…"

Woman: Rob, we're going to have to wrap this up, so…I don't know if you have too much more that you want to…

Seeman: Oh, go ahead, go ahead. You might want to rename it "more stress day and less stress day," because it's the average. The black line is really the average stress test for that person? Is it? Or no?

Stawski: No, no. I centered it so that it reflected a non-stressor day. Okay, so, I will try to breeze through these quite quickly, but I just wanted to show that the large-scale field studies of naturally occurring salivary cortisol rhythms are feasible. We can recover these rhythmic parameters that we would expect to find their strong evidence of variability within these diurnal rhythms across both individuals and across days, which is nice when you're trying to model differences of why are some people healthier than others, as well as why some days are healthier than others. And it's nice to use some of these last findings and actually find some corroboration with experimental evidence where we see that laboratory based stressor challenges shift cortisol levels in such a round where we can see that there is some association between experiences people have on a daily basis and their parameters. Their diurnal rhythm, of course, all fit well. And, then, finally, just some points I wanted to bring out, and we would love to get feedback on this and some perspective on this. When we model cortisol, as has been done by numerous people, we use natural log values. Log-transform cortisol, which, as far as we've found, has been done to satisfy distributional assumptions, but this was borne out of experimental studies trying to do group comparisons. However, we were log-transforming the values, and then modeling a rhythm, we're not picking the exact same thing. So, those are the issues with modeling diurnal rhythms with raw coritsol levels or natural log transformed cortisol levels. Of course, we are continuing to find ways to treat samples and detection of non-compliance, and this is…what we were getting at. Instead of necessarily excluding data, possibly including that this was a bad day, as a covariant in the model, which the graph I displayed with the good days and invalid days was that sort of model. So, it wasn't estimated in the two sorts of day separately. It was estimating an interaction. And then also, finally, is our approach in terms of our econometric correct? Maybe time of day, the actual fundamental time of day – 7AM to 7:30AM may be more important than simply taking a half hour difference for a given individual, so to what extent is diurnal rhythm of cortisol actually linked fundamental clock time instead of the individual's time? And, also, considering some other analytic approaches, which I know Arun and Teresa had been working on. It's not [inaudible] models where there are certain change points that occur at specific time points which are theoretically driven by where we would expect to see changes in the diurnal rhythm of cortisol, thirty minutes, and then, maybe, other things. So, that…that's it. Sorry I went over time.

Diane Lauderdale– Biomarkers of Sleep

Lauderdale: Some health effects of sleep duration are well known. People who sleep less are more likely to have accidents. In sleep laboratory studies, people with less sleep have decreased vigilance and poorer reflexes. That all seems pretty obvious too, although it is not always obvious to the persons themselves.

More recently, there's been intriguing work done on the memory, showing that if people do not get a reasonable amount of sleep, they don't retain information from the night before. Their problem solving abilities improve the next morning on a task they were introduced to the previous night if they slept well. That's all quite intriguing and speaks particularly to the general area of adolescence and sleep.

But the interest in epidemiology has really been bolstered by the evidence that there could be mechanisms that link lack of sleep to obesity and diabetes. This work really started at the University of Chicago in Eve Van Cauter's lab. In 1999, she published a paper in The Lancet which was based on a of couple dozen healthy young men. They found that if you only let them sleep four hours a night for a week, the hormones that are thought to affect appetite -- their diurnal patterns -- were affected in a way that made them feel hungrier (cortisol, lectin, grelin), and that has been confirmed in additional sleep lab studies. In a followup study, subjects received constant glucose amounts so that they received the same number of calories, and those studies revealed that subjects who slept less felt hungrier than the others. This is intriguing, the idea idea is that people who don't sleep as much are going to feel hungrier and are going to eat more. There is also another way there could be a connection: simply the more hours you're awake, the more opportunities you have to go to your refrigerator. Van Cauter's group also found that people's glucose metabolism was poorer if they got less sleep. Coupled with this last point is the general perception that people are sleeping less and less in the US.

I (and others) have found that part of what makes sleep so intriguing is that it appears that there are strong socioeconomic correlates to sleep. What we found in the study, which you'll learn more about in a few minutes, is that whites sleep more than blacks (data from CARDIA). People with more income do a better job of actually sleeping during the time they're trying to sleep. So, while they don't spend more time in bed, they get more sleep than people with less income. People with higher education similarly sleep more effectively compared to those with less education, and women sleep more than men – quite a bit more. This is fascinating because women make up most of the patients in sleep clinics. They're the ones that complain much more about insomnia and not getting enough sleep, and it's unclear whether this is because men and women are truly physiological different or men are just oblivious.

Usually, I laugh at people who show slides like this, tracking the number of journal articles on a topic over time, but I did something a bit more defensible. In Pub Med, this search is limited to the core clinical journals, so the actual volume of publications is being held generally constant here, and you see that, really around Van Cauter group paper there in '99, something is happening. The number of articles that have epidemiology somewhere in the indexing and sleep

the self-report kind of question, which a lot of studies do because it is a very low burden. There is the sleep log, which sleep physicians feel comfortable with as being a reasonable way of getting more accurate data, and it seems likely to be more accurate. What it doesn't tell you about -- unless you have supplementary questions – is whether there are waking periods in the middle of the sleep. And people are very bad at estimating how long it takes them to fall asleep and how long their waking periods are during the night. But still, this seems likely to be somewhat reasonable data. The gold standard for measuring sleep is polysomnography, which is what you do to diagnose breathing problems like sleep apnea in a sleep laboratory. There is actually a relatively recent home version of polysomnography. Polysomnography is the best way – really, the only way to get an apnea diagnosis. It's the only way to know the architecture of people's sleep, that includes things like how much REM sleep they have, or deep sleep. It's not a great way of measuring duration, though, because by the time you've got the monitors hooked up, you've got to think that that's going to affect your sleep, and it does because when people have done polysomnography over several nights, they see that there's a first night effect. So, the best studies have a habituation night, but it's quite expensive to do so. Even so, it really is not a natural circumstance, even in the home because, you have to schedule the technician to come and hook up the equipment, and it might not give you a realistic sense of how much people usually sleep.

So, this leaves wrist actigraphy, which has been added recently to several epidemiologic studies. A wrist actigraph looks like a really cheap watch, but in fact costs over $800+. It's an omnidirectional pedometer type instrument that records a measure of movement, and various models hold data for seven to fourteen days. You download it after someone has worn it, and you get something like this printout, which gives you a record of their movement. And when people are asleep, you get much less movement, but there is still some movement. If you actually see no movement, it means they took the watch off. Which happens -- you might not want to shower with the thing on, even though it is waterprrof. And, so, the software takes this and analyzes for total sleep duration and also some measures of fragmentation and efficiency, how much you're awake after sleep onset, and a bunch of related measures. I'm just going to focus on duration at this point.

So, this is an attractive way to measure sleep in an epidemiologic study. In a healthy population of adults, not the elderly who can stay really still when they're not sleeping in bed, or adolescents who might move more, especially if they have attention deficit disorder . But for healthy adults, you can do a very good job of figuring out how long people are sleeping. And in a sleep lab, when people have polysomnography and concurrent actigraphy, there's a correlation of over 0.9. Now, that's actually, which no one points out, also somewhat of a peculiar validation because it is undoubtedly true that the amount of movement people have while sleeping is different when they are hooked up to the polysomnography. So, you can't really do ideal validation here.

It's also unclear whether polysomnography is a perfect way to measure sleep duration either – the two methods key into sleep onset at slightly different phases. The actigraph kicks in and calls something "sleep" a little bit before the polysomnography does – usually by a couple of

minutes or more. Reading polysomnography for shallow sleep is not very reliable. It's quite reliable for deep sleep.

In this study, we are at the point where we've finished the data collection, and now we're finishing up the analysis. This was an ancillary study to Cardia. And our study asks whether objectively measured sleep predicts changes in weight, waist circumference, fasting glucose and insulin , estimated insulin sensitivity, resting blood pressure, and then also to look at how sleep characteristics vary by demographic, socioeconomic, and other factors.

CARDIA started in 1985 and recruited a study population that was half-white and half-black, half-men and half-women, higher and lower education. However, participants continued with their schooling after study enrollment, so the education level is generally high. Perhaps people who are good at staying in school are also good at staying in a cohort. And, so, for our study, we have a five year cohort nested within CARDIA, where the year 15 CARDIA interview and examination forms the baseline, and the year 20 represents the outcome. So, our outcome is change over five years. Our funding started in April 2003. We are working with Kiang Liu and others at Northwestern. Because of our funding date, we could not start measuring sleep any earlier thn we did. So, we have two sleep measurements a year apart for each of the participants in the study. And we only did this for budgetary reasons in one of the four CARDIA study sites. So this is just out of Chicago-Northwestern And not the other three CARDIA sites.

So, Cardia altogether had about 5000 people across the four sites, and it's a really high quality cohort study, but they did, of course, have some attrition. There are 814 non-pregnant participants in year 15 at Chicago. We invited them all to participate. 670, or 82% agreed to participate. We had something quite wonderful in CARDIA, which is, in year 15, they asked that self report question. "How much sleep do you usually get" and "do you have trouble sleeping". Using that, we were able to see something you don't get to observe often in epidemiologic studies. Regarding the 18% who didn't participate and the 82% who did participate, their own perception of their sleep was exactly the same, which is important to know. So, we can generalize to the entire population of persons who stay in a cohort study for twenty years.

For each of the two waves, we have three days of actigraphy. We asked people to wear the actigraph on Wednesday, Thursday, and Friday nights to capture two weekday nights and one weekend night. We also have self-reported sleep duration and some standard survey sleep instruments.

The biomarker setup here is that the actigraph is what I'm calling the "biomarker". I think this situation is similar to what a lot of studies have, which is that we have a self-report or "worse" measure for an entire cohort. We have a biomarker collected for a subset of the study for budgetary and logistic reasons. It's not a self-selected sample. And we have repeated biomarker measures because we collected three days in a row, but then we also have three days a year apart. This data setup, and the kind of the analytic issues they allow I think are common.

The analysis sort of bears on other things that could be done with the data, but a methodologically interesting question is how similar were our self reports to the actigraph measures And we're choosing to parameterize that by focusing the self-reports on two dimensions. One is the average difference between the self report and the measured sleep, and one is the correlation. And then we're also interested in whether different participant characteristics, like whether they are obese or not, or have high depression symptom count scores or not or have more education or less education, affect the average difference or correlation.

So, to examine that, this is basically our model. We are predicting the report of subjective sleep. Now, we have a better question here, which is from the Pittsburgh Sleep Quality Index. It asks people "in the past thirty days, how much sleep did you usually get on a weekday and on a weekend". It is a little more specfic than the questions that are just "how much sleep do you usually get at night". You would think you might get slightly more accurate data.

So, we're going to predict the subjective base as a function of the object assessment and subject error. We're basically averaging the differences, the intercept here, and the correlation is the square root in the model because there's nothing else in the model.

We have a problem, though. We have a self-report that's asking about the past month and our objective measure is only for three days. So, that is a sort of an error-prone measure of the objective measure. And because of that, we have an errors in covariate variables problem. We're going to do an errors in covariate regression to deal with this. What happens if you don't do that? If we just had a three day weighted (or unweighted) average of the objective measure, we would have a weakened correlation. If everybody sleeps exactly the same amount every day, we're golden. But that's not the case. So, if we ignore the variability in the right hand side and in our objective measure, we're going to underestimate the strength of the association with the subjective measure. In this case, we don't want to do that because we want to give the subjective measure, its best chance to show that it is a reasonably accurate measurement.

So, the error variance and the right hand side variables can be quantified with the reliability coefficient, which is the ratio of the variance of the true thirty day average (this is something we don't have) and the variance with the "noisy variable" (our three-day weighted average). If it were perfectly reliable, if they were just the same, that would be one. The more variability we've got with three day relative to the thirty day, the reliability is going to approach zero, and then measurement error becomes a really big problem. So, if we could estimate the reliability coefficient, we can correct for this and get rid of the attenuation bias in the regression model.

To estimate the reliability coefficient, the reliability is the true variance over the true variance plus the error variance, which, through algebra, just comes down to being the total variance minus the error variance over the total variance. So to estimate this from the actual data we have at hand, the total variance we estimate by taking the weighted average for each participant, and then we just look at the sample for the 670 samples – the sample variance (our estimate of the total variance). For our estimate of the error variance, we looked at the variance within each person for the three days, and then we basically pooled those and added them. That's the error variance. And so using that, we can calculate the reliability, which overall was about 0.75, but

we also stratified people again by males versus females, whites versus blacks, obese versus not, and the reliability did vary somewhat across those groups.

Karlamangla: And that formula does not use the thirty day average anywhere?

Lauderdale: The thirty day…we don't have it. We don't have a thirty day objective measure. We only have the three days. If we had the thirty day, this would much simpler.

So, the reliability coefficient yields an estimate of bias in the naïve or unadjusted $\hat{\beta}$, which is to say what would have happened if we did that regression and ignored the measurement error. We used the bias estimate to generate bias-corrected estimates of the association between the objective and subjective.

Schumm: Diane, does estimating the reliability that way make some assumption about sort of the process over time since…

Lauderdale: We are assuming that those three days are like a random sample of three days of impurity. That's an assumption we can't really test.

Schumm: I mean, is that reasonable given small data sets that may actually have much…or they don't exist.

Lauderdale: We do see a lot of day-to-day variation. This paper is in press now at Epidemiology, and we fortunately received a sophisticated statistical review from them. They were concerned that there might be some night-to-night correlation in three nights. Curiously, what they were concerned is that they would all be atypically high or atypically low, which would indeed be a worse problem for our data analysis. But that's not what I think in fact goes on. I think that there is day-to-day compensation in sleep; these people are all in their forties. People who sleep well one night actually sleep less well the next night, so it's not like you have a good sleep week and a bad sleep week unless perhaps if you are sick. So, we did try to examine how that correlation across nights would affect our estimates of similarity.

Woman: Or you're writing a grant.

Lauderdale: Writing a grant, yeah. I don't know how many people write grants in the real world! We did ask them when Northwestern called to pick a week that was a relatively typical week. They weren't traveling or anything.

So this is what we found.

Karlamangla: Diane, the reason I brought up the twenty-day thing was not because I thought you had a thirty-day objective measure. It's just that the problem arose over [unintelligible] because one was supposed to be over three days, one was supposed to be over thirty days, but

the reliability measure did not have the number thirty in there anywhere. I'm wondering if the three day were a four-day comparison, you would probably not have worried about reliability as much since one is a three day comparison, one is a thirty day. If it were a three day average, it was a thirty day average. The number thirty had to come in somewhere in that reliability figure. It didn't.

Phil: Well, that was this point is basically the assumption was that those three days you have are sort of "randomly" sampled from the thirty, and so you get the same thing had you had the full thirty days.

Woman: Well, but it's an assumption.

Woman: They're both averages, though. You don't care about duration, I think. They're both averages over three or averages over thirty.

Lauderdale: Right, the average is going to be on the same scale.

Karlamangla: So, but if it was three versus three, then you would not worry about reliability at all.

Lauderdale: That's correct. We actually did this on a nightly basis too, so then there was no issue. We asked people "how much sleep did you get last night" in the second wave of data collection.

Karlamangla: Right, I mean, so if it was three versus ten, you'd worry about reliability a bit more than, but not much as if it was three versus thirty.

Lauderdale: That's correct.

Karlamangla: So, somewhere, that ten or thirty number had to come in this reliability picture. It didn't. That's what bothered me. But maybe the answer is that "thirty" is being treated like infinity.

Lindau: No, three is being treated as a random sample of thirty. So it's generalizable to thirty.

Lauderdale: I can send you the paper.

Mendez de Leon: The other assumption is that if you're asked about your sleep in the last thirty days is that you're actually averaging your sleep which is a fairly big assumption, I think. How long do you really go back? Is it more than four or five days?

Lauderdale: It's a terribly difficult question to answer, and given how much night to night variability we found, it became even clearer why it's such a hard question to answer. You're asking people to be a mental calculator, literally to work it out.

So, the mean sleep duration in our entire cohort was only 6.1 hours, which got some press when it came out a couple of years ago. People reported about 0.8 more hours of sleep than they actually got, so, on average, people's habitual report was 48 minutes longer than our measured sleep duration. But, many of those stratifying factors were quite important. These are the ones where people did not overestimate by as much. The people who were closer to what we measured were the obese, the people with high depressive symptomotology, people with a high apnea risk (this information came from a standard questionnaire that asked about snoring and trouble breathing and daytime sleepiness). People with high apnea risk and people whose self-reported health was fair or poor did not overestimate their sleep by as much as the people in what you would see are in the "more healthy" states.

This is a fascinating problem because it means if you're looking at survey data, the correlation between sleep duration and any health outcome, part of the correlation is probably spurious. It might have something to do with sleep quality and that is an issue. But, in any event, in any data set, even if there aren't any associations between obesity and sleep duration, you will get a bit of a correlation because of this reporting difference. So, this is an issue.

Phil: Yeah, but, does it go in the opposite way? Because shorter sleep duration is associated with poorer health and greater health risk, then the reporting on the healthy might be suppressing that association rather than inducing it.

Woman: They're overreporting. They're reporting longer.

Lauderdale: The obese are reporting shorter sleep. And those with poorer health are reporting shorter sleep. The problem is that we just don't know what reality is. Everything we know is from population based studies – that is, the correlations between self-reports of sleep and these health characteristics. It's impossible to know how much of that is actually due to this reporting issue. People who feel draggy during the day don't overestimate their sleep by as much, it seems.

Our correlation was certainly significant. There is definitely information about objective sleep in people's self reports of sleep, but the correlation wasn't high. It was 0.47. And the correlation varied dramatically based on our different kinds of variables. So, basically, in a lot of our comparisons, where the group did worse, the correlation was 0.2 or 0.3, and for the group that did better, the correlation was 0.5 to 0.6. This was true for race. There was a low correlation for blacks, and a high moderate correlation for whites. It was true by income. It was true by education. It was also true by sleep efficiency, which makes sense. So, people who actually sleep more of the time that they're trying to sleep do a better job of estimating how much sleep they get. Fascinatingly, those who reported fair or poor self-rated health had absolutely no correlation between objective and subjective. The correlation was 0.06, and it was not statistically significant. Although it is a group of forty year olds, most of whom are healthy, so it was not a large group.

So, what this told us was the methodological question. Self-reported sleep is problematic as a measure in epidemiologic studies.

So, in our analysis, we used the measurement error model, which I think is something that is pretty standard. It could be much more widely used, but it's only really of interest if the actual value of the beta coefficient is important. So, we really wanted to know exactly the 48 minutes and the 0.47 because if you just want to know statistical significance, it's not going to make much difference at all. But if you are actually interested in the beta value, this is a reasonable thing to do when you've got repeated measures of a biomarker and different measures because you think it varies over time. So, this would be a sort of sensitivity check to understand about how much variation there is, rather than the standard, which is just to average them all or to pick the middle one or something. It's a little more sophisticated analytically.

Also, though, if the error prone variable is a confounder, it could do a poor job of confounder adjustment without the correction. So, your adjusted estimate of what you're actually interested in could be off. So, it would again be worth at least as a sensitivity check going through this procedure.

We'll have plenty of time to talk about the bigger issue. This is a pretty nice biomarker measure on a subset. There are researchers who are using the self-reported sleep measurement in the whole CARDIA study of 4000+ people, and can we do something better given that we have this biomarker measure from the subset.

Schumm: I was surprised not to see a recommendation on your previous slide over using sleep duration and sleep efficiency. So, what do you think the marginal [unintelligible] if one were to control for both. So, if I had 30-day measure, and I also include the efficiency, which does improve somewhat…

Lauderdale: Well, the sleep efficiency was coming from the actigraphy. That was the subset derived, but it was only coming from that. I mean, you could have sleep quality self-report, but that's another validation issue.

Okay, so this is the "in progress" part. This was challenging to think abouts. Can we do something better given what we've got on the sample? And this, I think, would be what we would want to do. We would want to impute sleep for the entire cohort rather than use the self-report individual sleep. We have a validation sample, and that's our subset for both measures. Unlike the previous analysis, in the validation sample, we would predict, the actigraph weighted average from both waves in that case, so they would be a year apart, which is attractive too. So, we would be predicting that on the basis of the self-report subjective measure. And we also know that we need interaction terms here because we just saw that there is variation between obese and the non obese, the educated and the less educated, etc. So, we could build a model that does the best job it can of predicting actual sleep based on self report and our other covariates. Then, we could plug that in instead of using the self report, and, (according to the biostatistician I am working with, Paul Rathouz), that's why it should give unbiased estimates But, we've got to come up with the standard errors, and we would have to bootstrap the

But, we've got to come up with the standard errors, and we would have to bootstrap the standard errors.

Karlamangla: So, did you create a regression model of the…

Lauderdale: We haven't done that yet. It's just something we've thought about, but it would end up being a methodologic paper, I think, rather than describing something that we found to have a strong association and presenting the self-report versus our subset versus the corrective self-report in the entire cohort.

Karlamangla: So, essentially, how to correct the self report.

Lauderdale: Seeing what you get from the imputation, which would be something closer to the subgroup. The subgroup, fortunately. is large enough to examine the association with the objective measure.

Woman: Two questions. What is the gold standard in this?

Lauderdale: Our gold standard was the weighted average of the actigraphy.

Woman: Right, so, if you used those data, would you still see some of those SES differentials.

Lauderdale: Oh, that was from the actigraphy. I think once we measured sleep with the actigraph, and that's what I was telling you about – the black/white and the…

Woman: Right, so all of that does exist in "real"…

Lauderdale: That really exists, and the race effect is gigantic at least…

Woman: So, the literature based on poor data is generally in the right direction?

Lauderdale: The SES correlates are weaker on the self-report than…

Woman: Which makes sense. So, it hasn't created a kind of differential that doesn't really exist. It's attenuated the real one is what you're seeing.

Lauderdale: Yes, that's what we've seen in CARDIA. Now, what we've also found something really funny, but I'm not sure exactly what forum to share it with anyone. So, they recollected self reported sleep in year 20, and part of the inducement to be in the actigraphy study was that we sent people a sleep report after each measurement. We also sent them $50 once they returned the actigraph because we really wanted those devices back. But, there's a potential for a learning effect. And what we found was that there was very little difference in people's self reported sleep in the larger Cardia cohort in 15-20, but people that had been in the sleep study had downed their estimate after seeing the actigraphy. And that was sensitive to education.

Woman: So, they could read the results…

Lauderdale: When they saw the report they might have thought, "My goodness! I thought I was sleeping 7 hours, and I'm only sleeping 6!" When you asked them again, they said 6.5.

Woman: Right, it makes sense.

Lauderdale: At this point, the well accepted idea that people are sleeping less and less is based on comparing answers people gave in, say, the American Cancer Society in '62 versus the American Cancer Society in '84. These are giant million-person volunteer cohorts, and comparing them wit the National Sleep Foundation's polls over the last few years.

Woman: But would that criticism also validate the Pittsburg sleep scales, or do you think that more elaborate type self-reported sleep questionnaires get at this better than the single kind of…

Lauderdale: The Pittsburg Sleep Quality Index is multidimensional. It's sort of getting at some of the things that have to do with apnea, some how you feel about your sleep, fatigue, and it seems to certainly be related to these things. It is, of course, so multidimensional that it's not really clear what all it reflects. Its questions about sleep duration are the ones we used, which is more precise about the weekdays, the weekends, and the actual time period. Some people recommend actually changing that time period to 7 or 14 days, thinking that that might get more accurate data. Maybe it does, maybe it doesn't, because in the second wave of data collection, we asked people on a sleep log to keep a sleep log, but they were asked to state how much actual sleep they got the previous night, and we compared that to the actigraphs, and it's only slightly better than the habitual sleep measure. So, people don't really know what happened the night before.

Part of the problem is that people have trouble with clock map. So, you get people who say they tried to fall asleep at 11, got up at 6, but got 8 hours of sleep. So, it's a fascinating mess. But sleep does seem to matter! We have some extraordinary findings, not for the hypothesis we actually originally set out with. There's a paper under review now where we found a massive effect of sleep duration on incident coronary artery calcification, which is fascinating. It's a clinical measure of aertherosclerotic disease, which is very well done in CARDIA.

Lindau: Well, what was the actual relation? Less sleep or more sleep? Because there's this new stuff that too much sleep, or people who sleep a lot might be sicker.

Lauderdale: The findings were that less sleep is bad. I think that the people that are saying that they get nine or ten hours of sleep, are telling you something about their life psychologically. It's not really telling you anything about their sleep because I suspect few really sleep that much unless there's some real physical problem.

Woman: Yeah, I was just thinking about the psychological benefit of thinking one slept more than one did. Does that have any role? When we had the sidebar earlier, we were talking about the stress that's related to thinking you're not sleeping enough, and that interrupts all of the

causal things that we're talking about.

Lauderdale: Yes, this is interesting. I don't know the answer to this. There are all kinds of interesting issues, and I didn't finish, but the reason that the general population surveys report an hour and a half less sleep than they did in the '60s does not necessarily mean that people are really sleeping an hour and half less. In the 60s, everyone was told they were supposed to get eight hours of sleep, and almost everybody answered that question "eight hours". It might not have much to do with what they were doing. It might or might not. And nowadays, people know that nobody's getting enough sleep. So they answer what they think is…I think this is like a different survey question. Like, "how much sleep do you think people are getting", not "how much sleep are you getting". This is potentially because people don't know how much sleep they themselves getting.

Thom: Well, sleep is also taking on a moral tone. It's sort of a testament to how hard you're working, or how busy you are.

Woman: Kind of back to your actigraphy data, I was wondering if you looked at the association between the sleep quality measures from the actigraphy to the subjectively rated hours of sleep. So, the number of night awakenings, sleep efficiency, late sleep latency, those types of variables.

Lauderdale: Right, so we created these stratifications based on higher versus lower efficiency, higher versus lower fragmentation, higher versus lower variability. People who had more than two hours difference in measures of sleep over the three nights versus people who didn't, and the only one of those that really mattered for the closeness of the objective/subjective relationship was efficiency. People with high sleep efficiency have a much better sense how much they're sleeping than people without.

Woman: I'm also interested in not just how that affects the accuracy, but the raw association between those sleep quality and the link with their perceived duration. So, you're predicting to the degree of accuracy rather than…

Lauderdale: I think it actually speaks to the same [unintelligible]

Woman: Is it? Maybe. Don't worry about it. It's subtly different, but don't worry about it.

Lauderdale: I mean, generally, people with worse sleep quality think they're getting less sleep. That is true. So, you're getting a perceived quality measure playing into the…

Woman: And if quality matters more than duration, then the self-report measure is better.

Lauderdale: It's not a particularly good measure of quality either, though. It's probably, again, significant but has an even lower correlation.

Woman: Right, that was the question. Is it a better correlation?

Lauderdale: This is an opportunity, to come up with something that is actually correlated with sleep quality.

Woman: Well, "do you feel rested is" one of the questions that's often used as a self-report of perceived sleep quality.

Lauderdale: Yes, but feeling rested has to do with both the quality and the duration. You can sleep beautifully, but if you only have five hours, because of a deadline…

McDade: But, in relation to health outcomes, it's an empirical question on whether self-reported or objective duration and self-reported efficiency versus actual efficiency are part of the outcomes that we're interested in. So, that should be your next study.

Lauderdale: Yes, we have been looking at that.

McDade: Can I ask a quick follow up about the comparisons of the objectives and subjective you did? Did you look at the distribution of sleep duration and how that might affect the variance as well as the relationship? So, for example, people who are overweight underreport their weight, people who are underweight overreport their weight, so the difference is not consistent across the distribution of values.

Lauderdale: Yes, but the issue here, though, is that everyone is sleeping less than people think they're supposed to sleep.

Woman: Well, Cardia is not self-report weight. They bring these guys…

McDade: No, no, no. I'm just using that as an example. There's a very easy way to do this that might be instructive. In the lab when we're validating a method, we plot on the x-axis the average of the score on the new method and the gold standard method. On the y-axis we plot the difference between the two. It's called a Bland-Altman blot. It's a very effective way to see if the distribution has shifted up or down and whether that difference is consistent across the full range of values. You might actually be able to see variation…

Lauderdale: We cut it in the publication because it was getting too confusing. We also looked at the calibration and how that varied across. Basically, people who report less sleep are more accurate than people who report more sleep because everybody's sleeping less than they think, so, yes, that's true. So if someone tells you that they're only sleeping five hours a night, they actually might be pretty accurate. If somebody tells you that they're nine hours, the might not be so accurate.

McDade: Yes, interesting. Yeah.

Lindau: So, the self-report and objective measured sleep are correlated. It's just that they're correlated differently depending on who you are and your characteristics?

Lauderdale: And also the correlation is not has high as people. I think epidemiologists were kind of assuming this was almost as good as asking people what their weight is, so you would get a correlation of 0.8 or 0.9.

Lindau: Right, so that's the important thing. So, we are dealing with self-reported tobacco use and salivary cotinine levels, and, we find that the two are very highly correlated. If you say you smoke, your cotinine levels are high. If you say you're exposed to smoke, they're higher. They're lower compared with smokers, but they're still high. So, those two things are very well correlated. We then look at a third thing, which is, in this case, HPV, human papillomavirus. It's highly correlated with self-reported smoking, but not significantly correlated with the cotinine values. So, is there an analog? Were there ways in which the self-reported sleep is correlated with an important health outcome, but the objective measure is not?

Lauderdale: The self reported smoking reflects some longer-term behavior variability that the more short term cotenine measure is not picking up? So, I don't think you can sort of say that it's spurious smoking data that people are…

Lindau: No, I don't think so, but what I'm trying to get at what is it if the two are relatively highly correlated, what is it that the self-report either in sleep or smoking is capturing that correlates with the health outcome?

Woman: Yes, is it actually just sort of a significance thing or a variability or is it truly the point estimates? Are they different? Because sometimes not really telling you something that's that different just because we're fixated on…

Lindau: Right, yeah.

Schumm: But, Stacy, you're talking about the difference between the full cotinine measure. You're not talking about an indicator of smoker versus non-smoker derived from the cotinine versus the self-report, right? So, in other words, couldn't it possibly be part of that there is some variability in cotinine with the smokers and within the non-smokers, and couldn't it be that that variability is not really associated with HPA or HPV outcome?

Lindau: Yes, but as I recall, we looked at it both ways. So, even though the cotinine distribution was very clearly divided into people who were never smokers, you know who the second-hand smokers and current primary smokers are. I believe that it was not significantly predictive of HPV status. I also don't think it was significant where self-reported smoking status was strongly correlated with current HPV infection.

Woman : So it might have been the second-hand smoke issue [unintelligible].

Lindau: It may be. I don’t know. I don’t want to derail this conversation, but I just wanted to know if there was an analogue to the sleep story.

Woman: We were supposed to be ending now anyway.

Man: Okay, thanks very much, Diane. That was great.

Steve Cole – Social Isolation and Gene Expression

Steve: …if you have variables, columns, subjects, rows, and look at the orientations of these things, we have about 20 rows and about 20,000 columns. So obviously, we're not going to be doing normal regression analysis on these things. There are strategies that biologists increasingly use to deal with this explosion of outcomes, but a lot of them are sort of strangely agnostic with respect to hypothesis. What I'm going to be talking about mostly today and the [inaudible] is how to take your biobehaviorable hypothesis, broadly speaking, and use them in the data analysis strategy to pull out gems. So, this is that huge plot I was talking about, and we have to look at each one of these rows. In the first one, we have about seven bony people from John Cacioppo's studies and about seven non-bony people for the socially integrated people, and each column is one of 22 or both, so [inaudible] that we actually assay. There are 200 or so are the ones that turned out to be reliably differentially expressed according to Paul's discovery rate of analysis, so already you can see that there are some big differences relative to traditional statistics. We're not going to be running around doing a 22,000 Bonferroni correction for the 20 people. That's clearly not going to happen. So, instead what we do is this false discovery rate-based analysis. Some of you might be familiar with that kind of thing, especially if you're doing [inaudible] because they do the false discovery analysis that are sample-wide Brown-Forsythe, but for the purposes of this talk, I'm actually going to skip over.

So, what we do is that we use a central false-discovery rate analysis, which basically says, "give me what everybody thinks they're getting when they get P less than 25," which means "give me this collection of genes, about 5% of which are actually not true, but at least I've got this quantitative sense of 'eh, 95% of these probably are real differences, they will probably show up again if I ran this study again.'"

This is a representation of the genes. As you can see, there are so many of them that the labels of the genes are actually really small, and that's, in some sense, a philosophical point for what we do. In general, the body doesn't do very much with any single gene. Most big changes in physiologic function are mediated by the activity of multiple genes simultaneously, so we are rarely looking for this one magic needle in the haystack. What we try to do is sort of read the tea leaves of genome-wide transcriptional profiles and ask "what kind of organized conspiracy is afoot in these data that is, changing the general regulatory regime of this cell," "how is this cell going to behave differently," "what pathways are driving those differences in behavior," and later on, we'll also talk about how differences in DNA sequence interact with these changes in gene expression profiles. But, for the first part of the talk, I'm really going to be focusing on social signal transduction and the process by which socioenvironmental influences get into the body, change in expression. Most of you guys are going to be pretty familiar with this. So, I won't bore you with my customary detail, but just suffice it to say, we'll be focusing, by and large, on things that come from the world around us, go through our central nervous system in the forms of perceptions of threat versus safety. That gets translated into changes in, for instance, sympathetic nervous system activity, or HPA axis activity, and it's really the product of those, what we call peripheral transmission systems – things like corticoids or catecholamine neural transmitters that are the ultimate proximal regulators of gene expression. They basically

bind to receptors and those receptors activate transcription factors, which activate the genes. They catalyze the expression of the genes. Most of our genes are actually silent in most of our individual cells. The majority of genes are involved in something else other than what this particular cell is doing. It might be turned on later if the cell changes its state right now, but usually, just because you've got DNA for a gene doesn't particularly imply that that gene is actually being expressed and therefore influencing the biology of your cells.

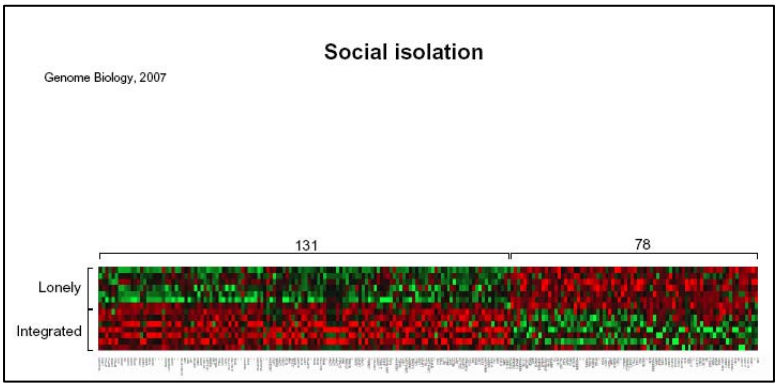
Karlamangla: Well, so, the expression varies from cell to cell in the body.

Steve: Very much so.

Karlamangla: So, the fact that you're looking at leukocytes, does that mean anything?

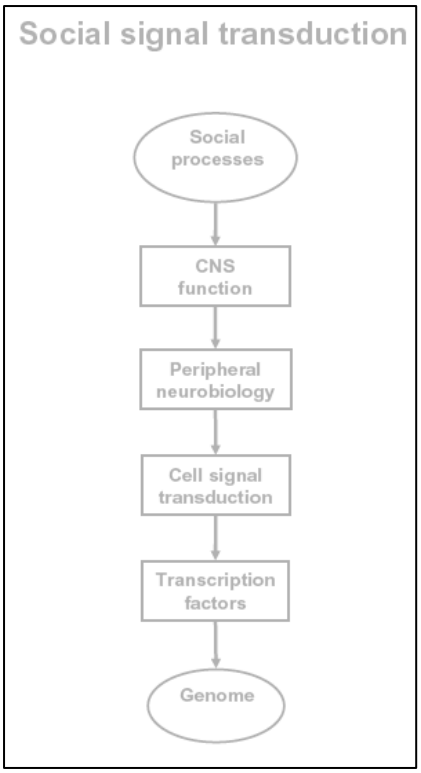
Steve: Yeah, that makes a big difference, and we'll talk about that a little bit later in terms of the generality of these effects. The take home point is that there are some common themes that arise in the body as a whole, and then of course, major distinctions relative to the specific activity this osteoblast is carrying out in the body as opposed to that neuron or something like that.

The problem is that we've got 22,000 of these genes here, and we're trying to basically analyze them in the context of these microarray data. The most efficient thing to do is to think of this in the way that the social scientists have used multi-indicator models, and that is to try to identify these conspiracies and measure the level of the conspiracy that's going on – treating individual genes as indicators of these underlying themes. This is going to be the overarching analytic strategy. The themes we're going to be looking at are themes related to the particular transcription factor driving this set of genes. For instance, here is one transcription factor that we mapped out as on that impinges upon some subset of these genes. These genes are distinct from another transcription factor that impacts perhaps a partially overlapping set, but also serves as a big complement of unique genes. These forty play genes and these three play transcription factors are a realistic representation of the magnitude of a search-based reduction that takes place when we move from individual gene analysis to theme-based analysis – things like specific transcription factors. There are other themes that are also important, which we will talk about in their relation to gene function. Is this gene product involved in transducing neurotransmitter information? Is it involved in regulating energy metabolism? Is it involved in regulating immune responses, etc. So, the organizing questions for the work that we do basically involves coming up with the substantive question and then applying this multiple indicator approach to the microarray data analysis in order to answer those questions. And the questions that we focus on are questions



like how broad are these effects? Are they causal? I'll give you a little bit of data on that in second. Which transcription control pathways mediate those effects and how do the genetic polymorphisms modulate them? In other words, how is it the case that these kinds of socioenvironment regulations of gene expression might take place differently in me because I have a particular set of DNA sequence at a particular site in my genome versus Arun's.

Here's an indication of causal effects where we can, instead of going off and experimenting on humans, experiment on adolescent rats, who we can deprive of social contact for a couple of months. In this particular example, we're taking out their brain and looking at gene expression. Again, you can see that in five or six rats that were socially isolated for about a month, there are about 1500 genes showing increased activity in this particular region of the brain. About 300 show decreased activity relative to animals that are spending their adolescence in their normal highly-sociable environment. So it actually turns out that 9% of the total genome-changing activity in the brain of these animals increases by more than about 50%.



The other take home point is that there are very large effect sizes at the RNA level especially. So it's not terrifically difficult to find completely non-overlapping distributions of gene expression profiles for at least some number of genes as a function of some things like losing a night of sleep.

Woman: Is that the hippocampus, or what region you're looking at?

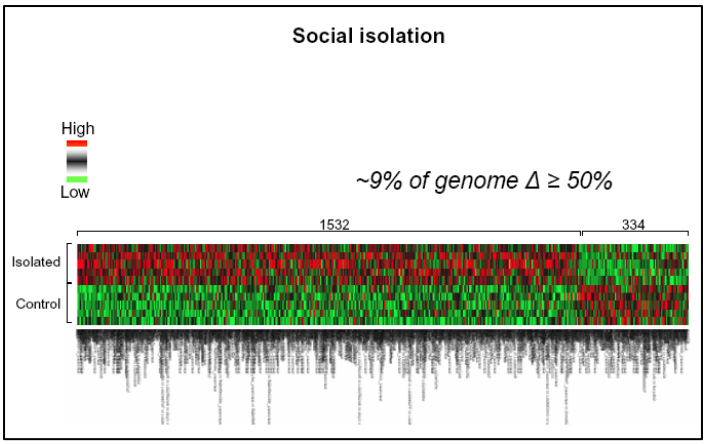
Steve: I believe this actually is not the hippocampus but, I'm embarrassed to admit that I cannot remember. I did this study about three years ago, so I just apologize in advance. I can find that out for you later.

Woman: …it's somewhere in the brain!

Steve: Yeah…brain, you know, does it really matter? Okay. So, the mechanism of the model that we use in guiding this whole thing is the notion that what we care about if we're trying to understand health itself is these profiles of RNA expression. DNA in the absence of some kind of expression typically isn't very influential, so as good biologists, we really have tried to follow the RNA more so than DNA stuff. The other thing that we're interested in doing is asking which transcription factor is actually landing on the promoter of this gene to turn on RNA expression. We don't think that all the several hundred of these transcription factor families that are, you know uniformly sensitive to socioenvironmental dynamics. We actually think that it's very likely that a relatively small number of these things are sensitive when you deprive animals of social contact because they don't fall over dead as a result. They don't even change most parameters of their physical health. Actually, most of their body and brain is still running largely the same way as in animals that are raised under normal social conditions, but

there are these identifiable focal differences. Therefore, the entire gene network is not falling apart. There's clearly some specificity here, and that's what we're trying to home in on.

So, one way of homing in on that specificity is to think about that cartoon I showed you, and especially the role of the transcription factors activating the genes. For example, is there a preponderance of certain kinds of transcription factors that seem to be driving the overall profile of gene expression? In other words, can we account for that differential gene expression profile in terms of the transcription factor binding motifs that are present in the promoters of those genes? The promoter actually determines whether a transcription factor can turn a gene on or off or not. The regulatory region upstream of the coding region of the gene that the transcription factor that binds to the transcription factor actually just acts like a flag. It says, "Hey, somebody come along and express this gene." There's this generic transcription machinery that recognizes that flag, plops down, and actually transcribes the gene.

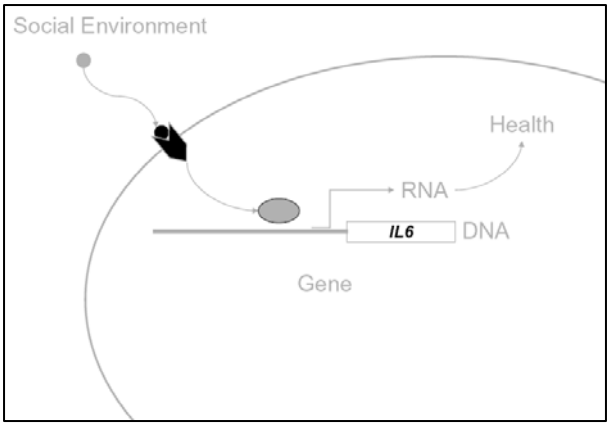


So, one of the ways that we can analyze this is to go through all of these promoters and ask "does this promoter have the DNA sequence that would allow, for example, the buka colicoid receptors sensitive to the HP axis and cortisol? Does this particular promoter have the DNA sequence that would allow this cortisol-induced transcription factor to bind there and catalyze gene expression?" The majority of the genes don't have it, and a tiny minority do, but it's certainly not the majority of genes. And we can do that for each of the other couple of hundred or so major transcription factor families.

So what I'm going to do right now is very quickly run you through some of the things that are emerging from those kinds of studies. We often see what you might call "adverse psychological circumstances" in the kinds of people who are under stress. These include people who are stressed out and people who you wouldn't necessarily automatically call stressed, like John Cassiopo's lonely people. They're probably lonely because they're trying to manage their stress levels. When we look at different samples of people under adversive psychological circumstances, we often find this thing that's depicted here. The genes that are relatively upregulated tend to have DNA sequences in their promoters that are sensitive to inflammatory transcription factors, particular NF κ B, one of the major central inflammatory mediators. So, to turn on an inflammatory response, the thing that your body does is that a receptor somehow recognizes tissue damage and activates NF κ B (or one of about 5-10 other proinflammatory transcription factors). We find an overrepresentation of these NF κ B response elements in the genes that are empirically upregulated in the leukocytes from those lonely people, and through biochemical studies, we've actually been able to show that we used to validate this kind of bioinformation approach. It's actually a pretty good inference. In general, when you find this overrepresentation of transcription factor binding, a direct assay of NF κ B will confirm that it

actually is increasing.

What we find underrepresented in, or to put it another way, overrepresented in promoters of the genes that are underexpressed are glucocorticoid response elements. This actually makes good sense to immunologists because they know that NF κ B and the glucocorticoid receptor are major antagonists in the body that cross-regulate one another. In fact, the body's chief physiologic way of maintaining control of inflammation is keeping NF κ B in check by tuning the glucocorticoid receptor. What we find is that there is no material difference in circulating glucocorticoid levels in these individuals who are lonely. So, clearly, all the way down from sort of the social environment to the brain to the peripheral circulation, that system does not actually have any marked changes. There's a tiny, tiny little attenuation of the HPA axis activity, but not in a good way to account for these big changes in transcription of glucocorticoid-mediated chains in these cells.



What we see is that somewhere between the glucocorticoid receptor and the actual transcription of genes, there is a disconnect in the lonely individuals. And that's actually consistent with a fairly well-developed literature and context of [inaudible], showing that the corticoid receptor can be desensitized by stress and, actually, also by inflammation to some extent. So what it looks like is happening is that the HPA axis is generally putting out the kind of signal that it should, but the genome is kind of adept to it because the glucocorticoid receptor isn't transducing that signal very efficiently in this case. We can actually drill down into the specifics of why that might be, but I think that's probably not going to be very exciting for this particular audience.

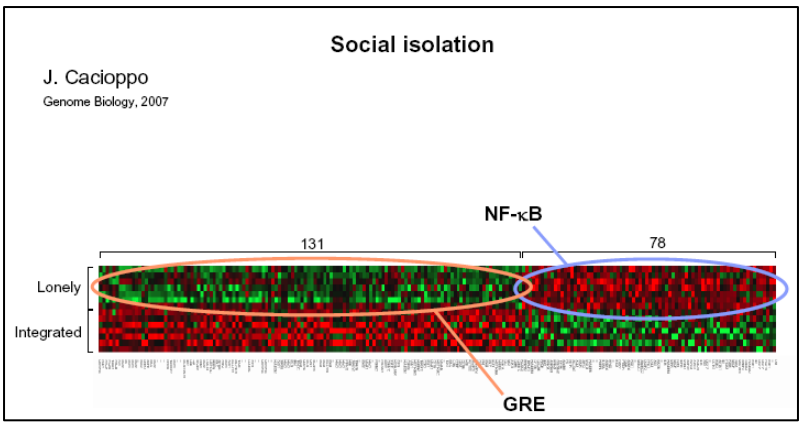
Lindau: Can I ask a question? So, is it that you looked at all the different classes of transcription factors and found that the inflammatory ones were most associated here or is it that you specifically looked at the inflammatory pathway, and you don't know what the association is with the other [inaudible]…

Steve: We did both. So, we [inaudible, voices overlapping]…that there would be this change in inflammatory biology that was actually motivated by epidemiology, showing that these conditions of social deprivation are associated with chronic diseases that often have inflammation as a major [inaudible], but we can't end with that hypothesis. So, we basically did a hypothesis testing statistical analysis using this kind of system. But, we also just looked at all whatever 198 other pathways that were there.

Lindau: Are there 198 classes of transcription factors? And did you find others that looked equally associated?

Steve: Yeah. I'll show you a couple in a couple of minutes.

As I mentioned before, these are not particularly distinctive to lonely individuals. So, for instance, we see very similar dynamics in another study with Greg Miller, where he looked at gene expression profiles from monocytes, which is one element of those circulating leukocytes that we were observing before. Pretty much, very few individual genes actually come up as differentially expressed in these two different samples. But, there is relatively consistent in this general theme of change in NF κ B mediated gene expression and changes in glucocorticoid receptor mediated gene expression. So, already, you can see the yield that comes from going out of the grass of these specific genes and up to these organizing themes. We see a lot more consistency at the level of organizing themes than we see at the level of the individual genes. There are some obvious reasons for that, as Arun mentioned. Different cells are going to be expressing different genes just as a function of their "purpose in life," but in addition, there's also some major statistical challenges because to run that false discovery rate analysis, even though it's better than, you know, a Bonferroni correction, it's still way, way too stringent in terms of if you were to do any kind of reasonable type 1 versus type 2 error tradeoff here, you're still way, way, way dialed up in favor of making a type 2 error. So, the big problem with replication studies isn't that the first result is right and a failure to replicate is wrong. It's the failure to replicate is so under-powered so that it's very, very hard to find an overlap that's applicable. But when you bump up to these organizing themes around inflammation or specific transcription factors, those actually have been strikingly reliable – more reliable than I ever would have anticipated. I've actually been shocked.



McDade: Steve, I have a quick question. On this slide, on the previous one, are you actually measuring NF κ B and other transcription factors, or you just know this to be associated with RNA.

Steve: We're measuring just that bioinformatic reversing prints from the RNA, however, we have done that stuff independently. So, once we get indication off the gene expression profiles, which are great for saying, "Hey, draw up 198 of these things and do your expensive confirmatory assays on those two." It's actually worked out pretty well.

Greg sees this same general profile when he looks at spouses of cancer patients. So, this is a very different kind of experience with the world than lonely people, but it's arguably still part of this general adversity. So, that's one general thing that we see. Also in Greg's case, we also did not see any marked difference in HPA axis output. Their coritsol trajectories over the course of the day seem to be pretty similar to a control population who didn't have spouses who have

cancer or had spouses who didn't have cancer.

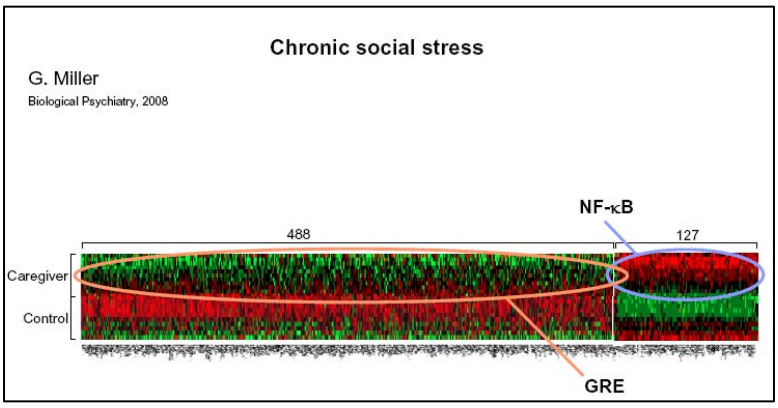
But at the level of transcription, clearly, the genome was acting as if they did have reductions in glucocorticoid signaling and increases in inflammatory signaling, presumably as a consequence of that glucocorticoid desensitization.

Lauderdale: What's the consequence in that the actual production has remained pretty much the same? What's the cost to the person?

Steve: Well, the cost is the genes run your health – not your cortisol. So, if your leukocytes are running around making more pro-inflammatory gene products, they're making more cytokines and causing more damage in the microenvironment of, possibly, a growing cancer cell. They're potentially contributing to the consolidation of an artheroscleroric plaque more efficiently. In other words, when we work from health, the most proximal determinant for health, and this is certainly the way of most certainly mainstream biomedical finders. The most proximal determinant of health is the RNA and cellular behaviors as a consequence of that. The glucocorticoid, to the extent that it doesn't play out to the RNA, is in some sense irrelevant to this process. But, you know, there's nuance to be had here.

Lauderdale: But it's irrelevant mostly because other things are being produced along with it. I have trouble seeing why it doesn't matter that somehow some compensation has gone on to keep those glucocorticoid levels as they should be.

Steve: It doesn't matter for the physical health if you believe that physical health is driven primarily by these RNA expression profiles. So, what you would like is the HPA axis to change its behavior in a way that regulates gene expression back into a healthy range. If that doesn't happen, either because the HPA axis doesn't put out the way it should or because the HPA axis is doing its job perfectly fine, the cell's not listening to it. In either case, you still have this proinflammatory gene expression.



Lindau: But if it's a feedback loop, how do you really disconnect what the brain is doing from the RNA regulation level of the cell?

Steve: The feedback happens actually at the level of the brain. There are actually two different feedbacks here. One is that the brain appears to not pay attention. So, one way the brain does this is that it just listens to how much cortisol is circulating the body and says, "Oh, got enough. I'm not going to make any more." The other way it does that is it listens to some of the

consequences of the cortisol, like the proinflammatory cytokine levels. So, it goes through and your brain will make more cortisol, hopefully to suppress that kind of thing. This is what happens in people who have relatively high levels of inflammation – acute levels of inflammation. If their glucocorticoid receptors are desensitized, they're not going to trim their proinflammatory cytokine output, and the brain may actually start trying to make more and more glucocorticoid. But the brain also listens to the glucocorticoid levels, so it may never be able to get a glucocorticoid level high to overcome the receptor desensitization. So, again, the focus specifically on the HPA axis totally makes sense, and that's definitely the right thing to do from a [inaudible] standpoint. But, ultimately, in terms of explaining their health risks, what you really want to know is what are the leukocytes doing. Are they out there in tissue making lots of cytokines or not? It's true that the brain will try and fix these problems, but if it can't fix the problem because the brain's receptors aren't exactly desensitized in the same way the lymphocyte receptors are, then you have this opportunity for the dysregulation of systemic inflammation. I'm not sure if I answered that cleanly enough. Does that basically make sense?

Lindau: I'm following your argument. Yes. I think I understood.

Steve: I don't actually require that I persuade you.

So, what are some of the other things that we see as themes coming out of these studies? One of them has to do with the activity of CREB/ATF transcription factors. These are [inaudible] of lots of different neurotransmitter receptors, and we see evidence of increased CREB/ATF transcription factor activity in John Cacioppo's study that involved socially isolated people, Greg Miller's study of people whose spouses are beset by cancer, and a study from Mike Kerwin where he does experimental sleep deprivation.

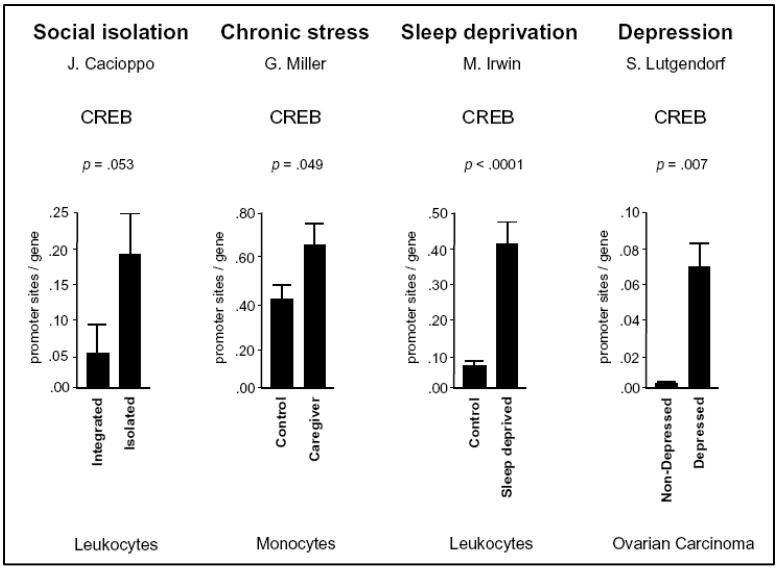
These are all results that come from taking blood out of somebody's arm, getting rid of the red blood cells, and looking at those gene expressing white blood cells to see what's going on. We can take a step away from that particular tissue site and go into, for instance, ovarian cancers, and find very much the same kind of dynamic in a completely different cell of origin. It's a totally different primordial cell type. It's a cell that's actively engaged in crazy behavior – like going crazy and basically killing the person. And, there, when we look at people who have high levels of depressive symptoms and low levels of social support, again, we find, as long as we have this relatively adversive psychological state, or negative psychological state, that, again, we see these elevated indications of transcription driven by CREB/ATF factor. So, this occurs in very different tissue sites. And, there's actually a couple of other examples that I'll show you later that have similar kinds of dynamics of the brain. Did someone have a question over here?

Lindau: Well, the last point I wonder about cause and effect. So, is it possible that the inflammation, some of which crosses the blood brain barrier, is the cause of the pressure?

Steve: In some cases, yes, in some cases, no. In Mike Irwin's study, he actually experimentally deprives people of sleep. So, we know that the causal effect goes from sleep deprivation. In Susan's study, we don't know that for sure. There isn't a feedback effect on the brain. There are ways to actually probe that sort of thing. The easiest way to answer that question, is to actually go into an animal model. We've done studies where we actually put human cancer cells into

mice that have no immune systems, and the human cancer cell grows like crazy. It seems to be showing exactly the same kinds of transcriptional dynamics that we see here. We've even mined in and measured the expression-specific gene products and found a lot of overlap. So, as long as we treat those observations as a portfolio of observations as opposed to looking at them in isolation, it's plausible that these are causal effects. But, at the same time, I'm not arguing that there's not some kind of feedback potentially going on as well because the ovarian cancer cells do make a ton of cytokines – in particular, IL6. That can be a pretty bad thing from a depression standpoint. So that's an absolutely reasonable hypothesis.

Another theme that we see coming up repeatedly are changes in what is called the innate-inner bio response that cells mount in particular to viral infections. One of the things we know about, especially shy sensitive introverted people is that they are at increased risk for a whole variety of viral infections. This is not such a problem for bacteria, not such a problem for some other diseases, but particularly viral infections are very sensitive to stress and also to these sort of temperamental things that structure your risks of experiencing stress. So when we look at, for example, Greg Miller's stressed people or John Cassiopo's disproportionally lonely individuals, we find that in both cases, reductions in indications of this type one innate anti-viral response is actually taking place, which is comprised of a key set of transcription factors called interferon response factors that seems to be under active immune sense. That actually makes sense in terms of explaining what the epidemiology is showing us, which is basically that these people experience accelerated disease progression – for instance, HIV infected individuals that have these kinds of psychosocial characteristics. Sheldon Cohen has done studies where he sprays rhinovirus up people's nose. These same kinds of phenotypes are associated with an increased vulnerability to infection and viral replication. So, this kind of thing provides some indication that there's a concrete molecular basis for these differences in post resistance to viral infections. The last thing I thought I'd show in this section is this bit from Mike Irwin where he experimentally deprived people of about 1/3 of a night's sleep. They woke them up at 2 or 3 in the morning just saying "if you can't sleep, just through the deepest part of your sleep there…"



Lauderdale: Actually, they're mostly missing REM then.

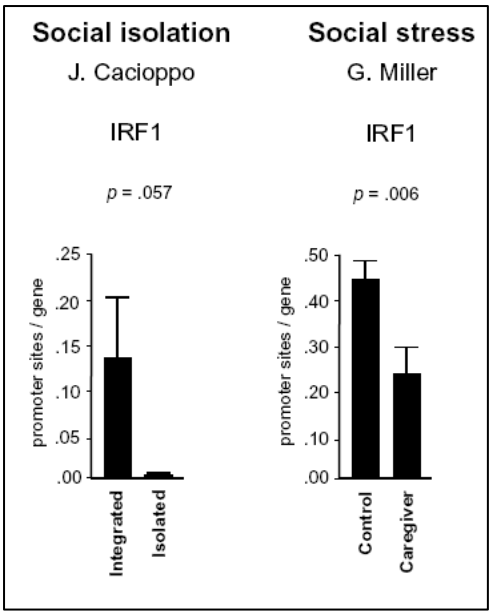
Steve: Yeah, that's probably right. He definitely wakes them up on the tail end. What we see there is that, again, from a smaller raft of genes from the leukocytes that's a fairly non-

overlapping distribution. Again, you'll notice that on a normal night, this pot of people were all agreeing on this gene product and the same five people in their partially sleep deprived nights were in the red for on all of these things.

One of the things that's important to note here is that when we talk about the psychosocial relationships to gene expression, it's not clear whether or not these big behavioral mediators aren't a key element of these kinds of things. Clearly, sleep deprivation and alterations in all kinds of other things could be the intermediate to these things.

Lauderdale: They could. In all those studies like that, it's a little ambiguous whether it's actually the sleep deprivation or the stress of being awake and then not allowed to sleep…

Steve: Right. Well, on some level it's hard to really do that one right for sure. I mean, you can side step the stress part by using biochemical mechanisms to deprive them of sleep, but then you've got the biochemical compounds, so you're absolutely right. It's very hard to parse those.



Lauderdale: There is also the case of somebody that naturally only sleeps 4-5 hours a night, whether that has anything to do with that.

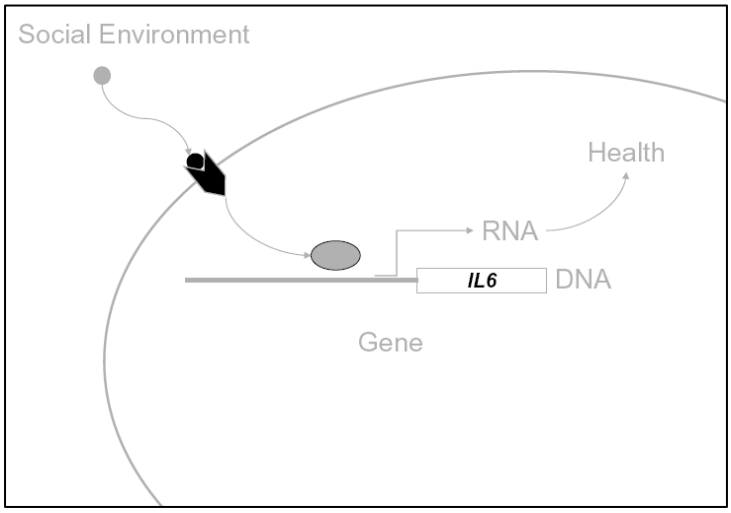
Steve: Right. That's a good question, and that's what I think we were asking about earlier. There are lots of people, especially as we age, who sleep quite a lot less than we used to. It's unclear whether they're at higher risk. But, on the other hand, we do know, for instance, that this inflammation dynamic increases with age. So it's not plausible, but it's somehow involved.

The last thing I'm going to talk about is how we look at these kinds of dynamics as moderated by your particular genetic sequence. In particular, identifiable occasions.

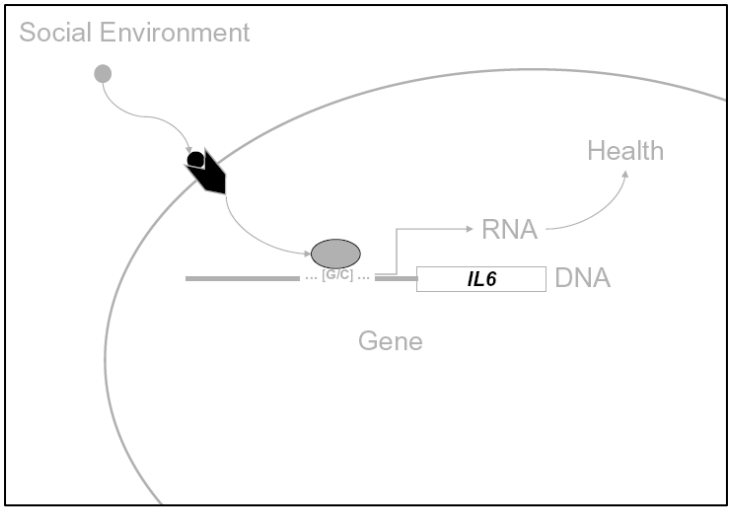
So we go back to our mechanistic cartoon where the way the social environment is influencing health is by turning on these transcription factors that catalyze the expression of RNA. And then we ask how DNA polymorphisms might modulate that [inaudible]. So, in other words, if I've got a G here and he's got a C there, does that somehow affect the ability of these social environmental stimuli to catalyze these gene expression profiles? And the way we do that is that you could go off and do these genome-wide association studies. They're very popular these days. You'd get 10 million single nucleotide polymorphisms, and you'd try and [inaudible]… obviously, you've got power issues there.

A better strategy, we think, is to go in there and actually come up with a mechanistic hypothesis about particular polymorphisms and how they would affect the biology that's taking place with these genes. I'll show you one way in which we can do that. It takes advantage of the tremendous infrastructure we have in terms of genome sequences and things like that. In this

example, we're looking at the IL6 gene. IL6 is a usual suspect in lots of problems with chronic inflammation and autoimmune diseases. We're looking at the promoter, and that's actually a segment located about 170 base pairs upstream of the transcription start site – that's the part where IL6 actually becomes coded and transcribed. This is the regulatory region of the gene where the transcription factors are supposed to bind. We've got representations of these 200 or so transcription factor family motifs that we can slide across these promoters and say, "Oh, I think that's a site for a glucocorticoid receptor" or "That looks like a site for the GATA family of transcription factors." And what I'm showing you is this site where we find that is 75% of Caucasians have a G 174 base pairs upstream of the transcription start site out of those that have the wild type gene. About 25% of northern European Caucasians have a C on at least one of their alleles of the IL6 gene. So, the wild type sequence is a pretty good match for the GATA transcription factor map. This says, you know, the probability of a, that's not actually a [inaudible], that's not a probability at all. That's a match on a 0 to 1 metric, but what we know is that this empirical turns, and we're going to validate this with biochemical methods. That's a pretty sure hit. We're definitely going to see a GATA transcription factor right there. If you have a C there, it's much harder for a GATA family transcription factor to actually bind to this site. So what we're talking about, basically, is a single nucleotide changing the capacity of GATA factors to transmit some kind of environment into gene expression. This basically makes sense. Your DNA is in some sense gating the ability of an environmental stimulus, and, in this case, possibly a socio-environmental stimulus that we know that amongst some of the candidates that come up in these exploratory analysis that we were discussing earlier include GATA family transcription factors. So, it's not an implausible hypothesis about one pathway by which this information could be conveyed into the proximity of the genome.



The inference that we're drawing here is that this one single nucleotide polymorphism is a place where we could see genetic diversity in how the environment changes gene expression profiles, and to convince ourselves that we're not just, you know, sort of fooling ourselves with all kinds of computational strategies. We can go off and do experiments in a cell culture where we model the process of being stressed by putting this actual promoter into cells, like fat cells or monocytic cells, or other kinds of cells that would normally express IL6. Then we dose them with more epinephrine, which is the relevant neurotransmitter released by sympathetic neurons in the region that, particularly in lymphoid organs, where we're trying to model here. So, when we put the wild type IL6 promoter in here without the promoter or the epinephrine, we get about a 9-10 fold increase in activity with that promoter. If we put the same promoter in, except it has a C there instead of a G – no other difference – everything else all the same. We see only about a 2 fold difference. So, what that's



telling us is that if you've got a G there, you really are more sensitive to a socioenvironmental signal that might arrive through norepinephrine or anything else that would [inaudible]. That's why we attribute this specifically to socio-environmental stimuli.

But, essentially, that sort of confirms our suspicion that this is at least one place in which genetic polymorphism might moderate the effects of transcription factors that could be, in turn, the handmaidens of environmental influences. Actually, I think from the MacArthur study way way back looking at the implications of these kinds of relationships for health outcomes in that study where we have a fairly nice set of biomarkers to assess the consequence of these dynamics, and good health outcomes is one of…

Lindau: What other cytokines did you look at besides IL6? Is it possible that this is relative for IL6, but other inflammatory markers have the opposite…

Steve: It's, in theory, possible. It's unlikely empirically because of "Mendalian randomization," as we call it. That means that in general, this polymorphism won't be offset by another one unless there is some kind of existing special selective pressure. That is not an implausible hypothesis. It's a totally reasonable question. But for most of our candidates, it doesn't look like that's the case. In other words, we compute what's called a Hardy-Weinberg equilibrium comparing IL6 genotypes with the other polymorphisms that we find, as well as promoters of other pro or anti-inflammatory genes. But, we don't see any evidence of that.

To back your question up one level of abstraction, though, our strategy is essentially to look at 10 million single nucleotide polymorphisms. Searching all of those things is going to be a statistical lost cause. So, we focus our attention on the ones that look like they might really matter from a functional standpoint. And then we still won't end up with a ton of these things. This is just my reminder that if we, for instance, look for polymorphisms that might gate the ability of a glucocorticoid receptor to turn on the gene or to lock the induction of that gene, we can find 12005 really really good predictions. Things that go from a 25 binding estimate to a 0.6-0.5 estimate. That looks like it's really powerful. Once we get those kinds of hypothesis, then we do exactly what you suggest, which is we basically skate through all of this and ask how are we going to prioritize this stuff? Most of these genes are genes we've never heard of, and we don't even technically know. If we do, we can make inferences from the gene sequence, but not that accurately.

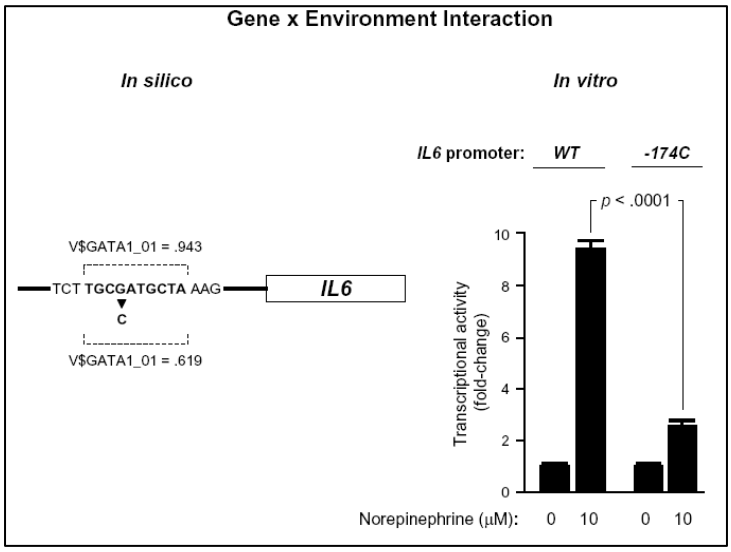
So, what we do is, to your point, we pick the ones that look like they're involved in inflammation or the control of inflammation – anti-inflammatory genes – to try to get at exactly

that kind of thing. They're going to be the main players for inflammation. Inflammation seems like a good place to work just because it's so mapped up in both the chronic health risks and things like macular degeneration. So, I won't claim that we're very far through that process, but that's kind of the general gist using sequence-based bioinformatics to use your hypothesis and prioritize things so that instead of running around and trying to run 10 million T tests on all of these different loci, you only focus on 10 or 1000 fold fewer of these hypothesis.

Lindau: So, the ones that you're choosing are the ones that others have discovered or others have documented a relationship for, or is there some other empirical basis for how you used the ones you think are most likely related.

What I'm getting to is that there's the problem, with non-genetic biological measures, that we keep using the same measures over and over because we know how to measure the thing and we've the assays that work. Is the same thing at play here?

Steve: Yes, to some extent. It doesn't apply at that first pass. In other words, at the first pass, we're completely unbiased. So, we'll end up with 1200 of these candidates for this transcription factor. We've got about 20 other transcription factors that look like good candidates, so now we're talking a lot of candidates. At that point, the first things we'll look at are things that we know from other information are important. But, very quickly, we can also bounce out of that world and use a set of inferences about the nature or function of genes that you don't know so much about. We prioritize them next for inspection. So, for instance, we can use information about the sequence of a particular gene, even if we know nothing about it's function, we've never actually built the protein and manipulated it, but "oh, that looks like a 7 transmembrane receptor" that's probably involved in turning things on and turning things off. So, we might prioritize that next. Ideally, we would like to get through this whole process. Even then, though, we have the risk that you pointed out. I mean, you might just be saying, "Okay, instead of looking where the light's shining, I'm looking 2 degrees outside of that," so what we've gone to recently is a sort of modified hybrid version, which we're not very far through right now. But, in fact, this is exactly what we're trying to do with John Cacioppo's studies – the blood spots that you guys collected and sent to us. We're basically saying that the genome-wide association study with its 10 million candidates is a complete disaster. There's no way you could do those statistics right, especially if you're running through a main effect hypothesis, when, in fact, almost all the time, when we find the main effects, it's with gene-



binding interactions. So, biomedical people in general are like, “Eeh, the environment, who cares, I just want to find a gene…” but the reality is that when you look who has the effect size, you know, [inaudible], it’s 10 times the size of the largest genetic risk factor for most of the chronic diseases out there. I think that makes it even harder to do with a large genome association study because we’re not including environmental variance in the analysis of this thing.

The high road we could take is to come up with this modeling of environmental interaction and then do an association study just the candidates based on that. So, when I say I’ve got 1205 glucocorticoid modifying significance, and a bunch of others from something like that, why don’t we just do a genome-wide association study? That’s what you might call mechanistically rationalized. That’s stronger. You focus your genome-wide study so if you can get 100 or 1000 fold greater power, and think especially beneficial if you’re doing the genome-wide analysis in terms of environmental influences and the potential of these SNPs to date them. We should empirically find things if we don’t know how to expect them. But, I don’t want to sound too sure about this. We know we’re going to miss things doing this. We know we’re also going to enrich faster than we’re going to [inaudible]

Woman: Do you know anything about the epigenetic modifications of these gene sequences, as in the availability of them for expression? Can you get at them with [inaudible] because that’s the step in between the…

Steve: Yeah, exactly. We don’t know much. We can make predictions about what, in principle, could be epigeneticlly modified, and we could use those for sensitivity analysis. We can say, “Let’s just methylate everything that looks like it could be methylated in here and then see what happens.” That, and I have to say, we haven’t really done very much, but the bigger challenge is the empirical question. It’s not so much, “Can we predict whether it’s methylated” because there’s a lot of things that could be or not. Our strategies for assaying the methylation, especially in a high [inaudible] way are not really up to this task right now. There are some strategies where you can pull out particular promoter sites with a methylation-specific antibody, and then hybridize them to a microarray – a CHiP on CHiP kind of a thing where the chromatin precipitation is done on methylation. But they don’t work very well right now. So, we’re still fairly far from being able to effectively incorporate that. Before moving to these kinds of genome-wide approaches, what we’re trying to do is actually work with people who are doing a much more focused analysis. For instance, we’re just starting some work with Greg Miller and Tom Woytz’s group at BBC who are all about methylation these days. What we’re trying to do there is to take people who are going to do methylation the right way. Instead of starting with transcription factors and then superimposing transcription, we’re going to start with their data on methylation and superimposed transcription factors to try and figure out whether you can clean up both stories at the same time. The general principle for data analysis is that the more systematic determinant side of the equation, the less noise you’ve got, the better your signal to noise ratio, the better you can sort of adjust all your estimates for confounding. We’re cautiously optimistic about that, but I can’t say that we’ve got a lot of empirical evidence that’s really going to work at this point. Seems like a great idea, though.

McDade: Would you have any indirect evidence for methylation or other epigenetic processes

of gene availability for transcription by looking at people who had the same genetic sequence for a gene and the same environmental influence, but very different transcription profiles.

Steve: Yeah, that's a pretty good way to do it. The problem is that that leaves us with a black box gate. Methylation could be part of that gate. It turns out that there's more that's part of it as well. So, methylation is one of a host of epigenetic modifications. It's an important one for sure because for the rest of you guys, epigenetic means everything except DNA sequence itself. Transcription factors get gated by epigenetic dynamics. Methylation is a situation where you basically methylate a particular base, and that transcription factor can't get on there. It's physically inhibited. There are other things that have the same effect. One of them is chromatin configuration, so DNA is normally packaged around these histones, which basically precludes access by transcription factors. That is another dynamic that we actually can't read off of the DNA itself because when we read DNA out of the sequence, the proteins fall off. So, there's a whole bunch of epigenetic things that could potentially be gaining transcription factor activity. We can get the totality of those through the reverse inference that you talked about in the exclusionary approach.

Woman: If you were to do the blood draw tomorrow rather than today or next week, do you know what the time course in the changes in expression are?

Steve: We actually know this really really well because we spent a lot of time optimizing our blood draw conditions to take advantage of these kinds of things. Most of this has been worked out in cell culture systems. When we try these things in humans, it looks pretty good. It seems to take about 5 minutes to convey that basic signal from immediately outside the cell into the transcription factors on the DNA. It takes between the peak of RNA change for most transcriptionally activate cells and about two hours after that. It can drag on out for a little while, and the tail of that peak might stretch out for several hours, but it takes awhile to get RNA starting to transcribe. I'd say that it takes half an hour to an hour to start up. It really piles ahead for another hour, and then it may keep going up if it has momentum, or it may plow down. The tail end of this is really unpredictable, though because there seem to be these intrinsic recursive systems for transcription control, where sometimes, if you start a gene up, it's going to keep itself going. It actually induces its own transcription factor that leads its own expression. So, this isn't a general law. You could have these things much more persistent. But, what we can say, at least in principle, is that it's possible that you could get these things on within an hour of two in a way that might be consequential, and the protein can hang around for a long time after that.

Woman: So, but the model for health outcomes is that there are going to be people that are chronically activated, over transcribing, and chronically ill after transcribing based on differences in the input signals and/or differences in the machinery. Right?

Steve: Right, right. And, in fact, I don't have an empirical basis for this right now other than cell culture systems. But I actually believe that the second thing is probably more important. You'll notice that when we did most of these studies, we didn't particularly mess with these

people. I mean, Mike Irwin, because he was lucky enough to be doing the experiment, messed with their sleep, but it's not like we actually hit them with some drug or some other kind of stimulus. This is because we're fairly pessimistic that these things will change things much. We think, at this point, and we were pleasantly surprised to see this, that probably the most important differences are exactly what you just said. It's what's happening to you under basal conditions every day, day in and day out for years and years and years. And because of this recursive non-linear control stuff, you can get yourself into a regime that propagates for a long time, and that, for the reasons you mentioned, really does strike me, at least, as the most likely contributor to these long-term inflammatory dynamics. We almost always look at baseline in human studies. In animal studies, like I said, we can do these parametric studies and say, that's what a perturbation would cause, but we've almost never done a human perturbation study besides Mike Irwin's sleep study. And there, we did the timing based on exactly the kind of facts I talked about. We think it's going to take about two hours for this to peak out. We got them at the point where they would have awakened had they not been deprived of these two hours of sleep, reasoning that that about 2-3 hour period should be about the right amount of time to see the peak RNA change. Any other questions?

Lindau: Has any one looked at these questions in prematurity or in relation to, for example, differential placental expression in low birth weight versus normal birth weight babies?

Steve: I do not know, to tell you the truth. I have no information one way or the other by and large.

Lindau: I guess I just wondered to what degree this was the result of evolution. This polymorphism that keeps getting passed on but doesn't seem to be advantageous.

Steve: Actually, it may be advantageous. We were happy to see this polymorphism pop up in our search because it's actually the best study of human polymorphism in terms of disease risk. For lots of chronic diseases, it seems to be deleterious. If you want IL6 around, it's great. If you don't want IL6 around, especially if you don't want IL6 around for years and years and years, which in general, you don't, then it's a problem. But there are certain things where, actually, it seems to be preventative. It's possible that there was a selective pressure for this someplace in northern Europe. The reason we keep saying northern Europe is if you look in Asian or African populations, you actually almost never see that thing. So, there's something weird there. You're absolutely right, it would not be conserved if it weren't somehow advantageous. It would wash out pretty quickly. And it sits right in the middle of a transcription factor binding site, and if you compare genetic sequences in humans and mice, you see that it's just total noise except where you get your transcription factor binding sites and the coding regions of genes. It's not just the gene product that's highly conserved, but within the promoter region, we see junk junk junk, and then here's the NF κ B site, which is pretty well preserved in each species. So, the fact that this would be carried along there means that it's part of that conservation process. But whether it's being carried by itself or because you cannot generally mess with this footprint without losing good control of IL6, we're not totally sure. So, what's the nature of the pressure there? I don't know that anybody has much good speculation about that. Also, the people working this stuff are, you know, surprisingly empirical. We don't like to think in terms of theory that much, so…[inaudible]

Chris: Just a footnote. I mean, there are notions of [inaudible], which has a role early on and therefore would be reverse selectively by [inaudible]. I mean, you could make an argument there.

Steve: Yeah, then the only question would be why we don't have this in African or Asian populations, but, absolutely, it's not hard to come up with a hypothesis. I just don't know if anybody has actually it down.

Woman: Well, this looks like the right breaking point for us. Thank you very much.

The Future of the Biomarker Conference

The participants had a detailed discussion about the future direction of the CCBAR annual conference and strategies for linking biosocial and biobehavioral survey researchers. For these details, contact Dr. Stacy Lindau at slindau@babies.bsd.uchicago.edu.