UNITED STATES OF AMERICA

* * * * *

FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

* * * * *

PERIPHERAL AND CENTRAL NERVOUS SYSTEM
ADVISORY COMMITTEE

* * * * *

MONDAY,
NOVEMBER 18, 2002

The Advisory Committee met in the Grand Ballroom of the Holiday Inn, 2 Montgomery Village Avenue, Gaithersburg, Maryland at 8:00 a.m., Claudia Kawas, M.D., Acting Chair, presiding.

MEMBERS PRESENT:

CLAUDIA KAWAS, M.D., Acting Chair
THOMAS H. PEREZ, M.P.H., Executive Secretary
ELLA P. LACEY, Ph.D., Consumer Representative
MICHAEL GRUNDMAN, M.D., M.P.H.
RICHARD D. PENN, M.D.
GERALD VAN BELLE, Ph.D.
HOWARD L. WEINER, M.D.
JERRY S. WOLINSKY, M.D.

MEDICAL IMAGING CONSULTANTS PRESENT:

CRAIG BEAM, M.D.
LEE C. CHIU, M.D.
MARK FOGEL, M.D.
HYUN KWON KIM, Ph.D.
JAMES M. PROVENZALE, M.D.
RUTH RAMSEY, M.D.

MEDICAL IMAGING CONSULTANTS PRESENT: (cont.)

GREGORY SORENSEN, M.D.
WALTER WOLF, Ph.D.

ALSO PRESENT:

RUSSELL KATZ, M.D., Director, Neuropharmacological Drug Products
PATRICIA LOVE, M.D., Director, Medical Imaging and Radiopharmaceutical Drug Products
ARMANDO OLIVA, M.D., Division of Neuropharmacological Drugs
ROBERT TEMPLE, M.D., Director, Office of Drug Evaluation I

I-N-D-E-X

P-R-O-C-E-E-D-I-N-G-S

8:00 a.m.

CHAIRPERSON KAWAS: If everyone can find a seat so we can begin.

Good morning. And welcome to the November 18, 2002 meeting of the Peripheral and Central Nervous System Drugs Advisory Committee of the FDA.

My name is Claudia Kawas, and the topic for today's meeting is the role of brain imaging as an outcome measure in Phase III drug trials in Alzheimer's Disease.

And we'd like to start by introducing the people who are sitting around the table, so perhaps we can start with Dr. Katz.

DR. KATZ: Russ Katz, Neuropharm Drugs, FDA.

DR. LOVE: Patricia Love, Division of Medical Imaging, FDA.

DR. OLIVA: Armando Oliva, team leader, Division of Neuropharm Drugs, FDA.

DR. FOGEL: Mark Fogel, Medical Imaging, Children Hospital of Philadelphia.

DR. VAN BELLE: Gerald Van Belle, Department of Biostatistics from the University of Washington.

DR. PENN: Richard Penn, Professor of Neurosurgery at the University of Chicago.

EXECUTIVE SECRETARY PEREZ: Tom Perez, Executive Secretary to this meeting.

DR. GRUNDMAN: Michael Grundman, University of California, San Diego.

DR. WOLINSKY: Jerry Wolinsky, neurology, University of Texas at Houston.

DR. CHIU: Lee Chiu, M.D., MI Imaging Director, California.

DR. RAMSEY: Ruth Ramsey, neuro-radiology and Professor of Radiology at the University of Illinois.

DR. BEAM: Just in time. Craig Beam, Biostatistician, University of South Florida, Moffit Cancer Center.

DR. WOLF: Walter Wolf, Professor of Pharmaceutical Sciences and Director Pharmakinetic Imaging Program, University of Southern California.

DR. KIM: Hyun Kim, Cal State University, Los Angeles. Chemistry ? Biochemistry professor.

CHAIRPERSON KAWAS: We also have our invited speakers sitting off to the left, and perhaps we can start with introductions there with Dr. Mike Hughes.

DR. HUGHES: I'm Michael Hughes, I'm a Professor of biostatistics at Harvard University.

DR. FOX: I'm Nick Fox, senior fellow at the University College London in London.

DR. De CARLI: Charles De Carli, neurologist, University of California at Davis.

DR. WEINER: Michael Weiner at the VA Hospital and the University of California, San Francisco.

DR. CHARLES: Cecil Charles, Duke Image Analysis Laboratory, Duke University.

DR. DORAISWAMY: Murali Doraiswamy, I'm a psychiatrist at Duke University.

DR. JAGUST: Bill Jagust, neurologist, University of California at Davis.

DR. SMALL: Gary Small, psychiatrist,

University of California at Los Angeles.

DR. JACK: Clifford Jack, radiology, Mayo Clinic in Minnesota.

CHAIRPERSON KAWAS: Thank you.

We'll now have the conflict of interest statement.

MS. TURNER: Good morning. My name is Tara Turner, I'm the backup Executive Secretary. I'm filling in in the absence of Tom Perez' voice this morning.

The following announcement addresses the issue of conflict of interest with respect to this meeting and is made a part of the record to preclude even the appearance of such at this meeting.

The topic of today's meeting is an issue of broad applicability. Unlike issues before a committee in which a particular product is discussed, issues of broader applicability involve many industrial sponsors and academic institutions.

All special Government employees have been screened for their financial interests as they may apply to the general topic at hand. Because they have

reported interests in pharmaceutical companies, the Food and Drug Administration has granted general matters waivers to the following SGEs which permits them to participate in today's discussions: Dr. Michael Grundman, Dr. Claudia Kawas, Dr. Richard Penn, Dr. Gerald van Belle, Dr. Jerry Wolinsky and Dr. Howard Weiner.

A copy of the waiver statements may be obtained by submitting a written request to the Agency's Freedom of Information Office, Room 12A-30 of the Parklawn Building.

Because general topics impact so many institutions, it is not prudent to recite all potential conflicts of interest as they apply to each member and consultant.

FDA acknowledges that there may be potential conflicts of interest, but because of the general nature of the discussion before the committee these potential conflicts are mitigated.

With respect to FDA's invited guests, Dr. P. Murali Doraiswamy, Dr. Michael Weiner, Dr. Nick Fox, Dr. Clifford Jack, Dr. H. Cecil Charles, and Dr.

Gary Small have reported interests which we believe should be made public to allow the participants to objectively evaluate their comments.

Dr. Doraiswamy attended a consultants meeting for Berlex several years ago, has received research grants and/or honoraria from Pfizer, Novartis, Eisai, Janssen, Merck, Forest, David, Elan, Organon, GlaxoSmithKline, Wyeth, and Lilly over the past five years. He has also received grants from the NIH, NARSAD and the American Federation for Aging Research.

Dr. Weiner has consulted for Pfizer, Aventis, Merck, Synarc and Novartis.

Dr. Fox has received consultancy fees or honoraria for lectures from Novartis, Janssen, Elan, Pfizer, Searle, Lundbeck and Pharmacia. His research has a collaborative research grant from GlaxoSmithKline and has been contracted to provide image analysis for Novartis, Janssen and Elan/Wyeth.

Dr. Jack has provided advice to Pfizer and Pharmacia regarding the use of MRI as a biomarker of disease progression drug trials in Alzheimer's

Disease.

Dr. Charles has a professional relationship with Duke University Medical Center's Brain Imaging Analysis Center and the Center for Advanced MR Development.

Dr. Small is a scientific advisor to CTI and Amersham and has an involvement in a pending NDA for FDG-PET in Alzheimer's Disease.

In the event that the discussions involve any other products or firms not already on the agenda for which FDA participants have a financial interest, the participants' involvement and their exclusion will be noted for the record.

With respect to all other participants, we ask in the interest of fairness that they address any current or previous financial involvement with any firm whose product they may wish to comment upon.

Thank you.

CHAIRPERSON KAWAS: Thank you.

I'd now like to turn the floor over to Dr. Patricia Love, Director of Medical Imaging and Radiopharmaceutical Drug Products.

DR. LOVE: Thank you.

Good morning, Dr. Kawas, all members assembled of the Advisory Committee, all medical imaging consultants, all invited guests. Thank you very much for coming. This is certainly going to be an exciting day and in a few moments, Dr. Katz is going to speak with you about the activities planned. But before then, let me just briefly address some issues about the status of the Medical Imaging Drug Advisory Committee.

Several of you received a letter over the last few days about the fact that the Medical Imaging or MIDAC Committee is no longer going to continue to exist as a standing entity. The basis for is certainly varied and involves several different aspects, but key among them is the fact that the agency is only allowed to have 12 advisory committees existing at any one particular time.

So, with need to add new committees, this is one of the committees that will no longer be in existence. But that does not mean -- and let me please reassure you that that does not mean that we feel that

there is any less need to seek input from advisors and consultants on this matter. And what you see before you today is a model that we will be using to seek your input and counsel. It's a combination of an advisory committee, a standing committee with invited guests, obviously, and that's one mode that we can use.

Another option we have is to potentially form an imaging subcommittee. If we do that, it would be a subcommittee of a standing committee. Before making that decision, however, because as you know there are products that are being transferred from the Center for Biologics to Drugs, we are waiting to determine exactly which products and what types of areas will be transferred before we make a decision on whether or not to form a subcommittee of a standing committee, and which committee that would be.

And, of course, the third option is the option we've been using several times over the last few years, and that's the pubic forums and workshops that we use for PET and positron emission and tomography issues, radiopharmaceutical issues that

stemmed from the Food and Drug Modernization Act of 1997. That type of venue allows us to have much more interactive dialogue with both the advisors and consultants as well as the public.

So we will continue to use all the methodologies. We certainly as an agency recognize and value the importance of imaging and this rapidly advancing technology, its relevance to diagnoses and to treatment. We will continue to move forward it in that area.

In the meantime, if you do have questions, please you may forward them to me directly or you may forward them to Linda Skladany, Assistant Commissioner for External Affairs.

Thank you.

CHAIRPERSON KAWAS: Thank you, Dr. Love.

Now for the FDA overview of issues, Dr. Russell Katz, Director of Neuropharmacological Drug Products.

DR. KATZ: Thanks, Dr. Kawas.

And I'd like to welcome you all here this morning. I'd especially like to welcome our medical

imaging consultants. You've just heard about their committee, or their old committee. And in particular I'd like to welcome our invited experts. Our invited experts will present to the Committee the state of the art of various brain imaging modalities that will form the basis for much of what we talk about today. So I want to thank you all for coming today and for helping us address what we believe to be a very important issue in the future, development of drugs to treat patients with Alzheimer's Disease.

Finally, also let me welcome Tom Perez, who?s filling in and has graciously agreed to fill in at the last minute as the Executive Secretary for today's meeting. So thanks very much, Tom.

As you know, today we are asking your advice on an issue that's become of considerable interest to manufacturers of treatments for patients with Alzheimer's Disease and a matter of interest to many other parties as well. And namely, that's whether or not we should rely on a drug's effect on a surrogate marker to support to support the marketing of a treatment for patients with Alzheimer's Disease.

Before I go on much more, let me just say what a surrogate marker is. There are many definitions available, as I'm sure you know, about what a surrogate marker is. And I thought since there were many available, I would take one offered by my boss.

I notice that I've actually neglected to attribute this to him. I thought maybe I could get it done before he came this morning. But he's here now, so I'll have to apologize.

This is from an article that Bob Temple wrote in 1995, and basically it says a surrogate marker and point of a clinical trial is a laboratory measurement or a physical sign used as a substitute for a clinically meaningful end point that measures directly how a patient feels, functions or survives. Changes induced by therapy in a surrogate end point are expected to reflect changes in a clinically meaningful end point.

Another definition that someone gave us of surrogate markers is something that you measure instead of thing you actually care about.

As these definitions imply, approval of a drug on the basis of an effect on a surrogate marker presupposes no requirement for a demonstration of a direct effect on a clinical outcome. And you'll recognize that that's unusual.

Obviously, the vast majority of drugs are approved on the basis of a showing of an effect on a clinically valid or face valid measure of how the patient is doing, whether it's objective or subjective. And, in fact, in our division all drugs have been approved on the basis of a finding on a clinical outcome, although on rare occasion a surrogate marker has been found to be supportive in some of the studies.

In fact, as you know, there are currently four treatments approved for Alzheimer's Disease and all have been approved on the basis of an effect on clinical outcomes, namely cognitive measures and global measures, as you know. But now we're being asked and we're asking you to consider approving treatments for Alzheimer's Disease on the basis of an effect on a surrogate marker.

Why would rely on the effect on a surrogate marker instead of a clinical endpoint? Usually two reasons are given. One is that because these measures are fairly sensitive, one could reduce the sample size necessary to show an effect and that, of course, makes studies cheaper and more manageable.

A second aspect of surrogate approvals that is often touted as being useful is that because surrogates often look at outcomes that may be very latent in the path of disease, mortality, for example, and it might be very difficult to study mortality directly, the use of a surrogate could decrease the sample size and actually make the study actually practical and more tractable.

And, of course, the Agency has a long history of approving drugs on the basis of effects on surrogates. For example, the obvious examples, are anti-hypertensives, drugs which are approved on the basis of a showing of a decrease in blood pressure, which is a laboratory measurement and not a clinical outcome in the sense of how the patient feels. Similarly cholesterol-lowering agents, as the name

implies, those drugs are approved on the basis of a showing of a decrease in serum cholesterol and not on any specific clinical outcome.

Here, though, these effects on these surrogates has presumably been shown to correlate with actually a clinical outcome of interests, for example, decreased cardiovascular outcomes or events. And in that sense these surrogates can be considered validated; that is to say an effect on the surrogate has shown to predict an effect on a clinical outcome of interest.

But in addition to approving drugs on the basis of findings on validated surrogates, since 1992 the agency has had the explicit authority to approve drugs on the basis of effects of surrogates that have not been validated but only that have been reasonably likely to predict the clinical effect of interest. And I have a definition. This is actually the language from the so-called accelerated approval regulations, again adopted in 1992. And I'll just read them.

It says that "The FDA may grant marketing approval for a new drug product on the basis of

adequate and well- controlled clinical trials establishing that the drug product has its effects on the surrogate endpoint that is reasonably likely, based on epidemiological therapeutic, pathophysiological or other evidence, to predict clinical benefit."

It's important to note that the regulations anticipated that these approvals would occur only for treatments for life-threatening or serious diseases for which there is no other available treatments.

In addition, the regulations also state that, ultimately, these surrogates would have to be validated, that is to say that the sponsor would have to demonstrate usually after a drug was approved or invariably after the drug was improved, that in fact there was a correlation with the clinical outcome of interest. And, in fact, if that couldn't be shown or if a sponsor didn't engage in that sort of attempt to valid the surrogate, the drug could be removed from the market more easily than other sorts of drugs.

And, in fact, in 1997 this essentially

same standard was introduced into the Federal Food, Drug and Cosmetic Act under what's called the fast track provisions. So this has been in the regulation since '92 and in the Act, the statute, since 1997.

I don't want to go into very much detail into the nature of validation of a surrogate. Dr. Hughes will talk about that, I believe, in a little while. Let me just point out that validating a surrogate is a complicated matter, and it ordinarily involves essentially a complete understanding of all the effects of a drug, positive and negative, as well as a detailed understanding of the path of physiology and the biology of the condition being treated. And as you'll also recognize, we usually don't have complete information on any of those matters.

So while the regulations permit the Agency to approve a drug on the basis of an effect on a surrogate that is unvalidated but only reasonably likely to predict clinical benefit, relying on the effects of a drug on an effect on an unvalidated surrogate is potentially problematic. These surrogates, and particularly the surrogates you'll

hear about today, imaging modality and Alzheimer's Disease, correlate very well with the untreated condition. In other words, as the Alzheimer's Disease gets worse, we see that the imaging modality gets -- modalities, all of them get worse in a highly correlated way, but it's not immediately obvious that a drug- induced effect on that surrogate necessarily translates into a clinical benefit that we want to see.

In fact, there are many examples in medicine where a beneficial effect has been seen on a candidate surrogate, but in fact the clinical effect of interest has not been shown. And a number of these examples are explained in various of the publications that you have in your briefing book.

Now, in the case of putative treatments for Alzheimer's Disease, actually we're not being asked by sponsors to rely on effects on surrogates for the documentation of symptomatic treatments. As I said, the four treatments that are approved, have all been approved on the basis of symptomatic effects and typical study designs are fairly good at picking up or

at least capable by design of picking up symptomatic treatment effects.

What sponsors are generally proposing when they ask us to rely on surrogates is a showing that the drug has an effect on the underlying program or path of physiology of Alzheimer's Disease. As I say, typically the study designs that are used now to look at symptomatic treatments are not capable of documenting such an effect on the underlying progression. There are clinical trial designs that are capable of demonstrating this effect, but those trials are very difficult to do. They involve or would involve large numbers of patients and would take long periods of time. So in that context, relying on a surrogate to document progression is very attractive.

In addition, the other reason that imaging modalities in particular appear to be attractive for this purpose is that they purport to give us a window into actually looking at the pathology. And so it seems reasonable to conclude that any effect that one would see on these imaging modalities in a beneficial way from the drug would necessarily translate into a

clinical benefit that we'd like to see. I would just caution that it isn't necessarily the case. It's a complicated matter, as I said before. One would have to at least understand what you're looking at in the imaging modality, first of all, in terms of the pathology and then there are at least three considerations that would have to be taken into account before we decided that an effect seen on the modality by the drug actually would translate into a clinical benefit.

And one is that one would have to ensure that there's no interaction between the drug and the test system itself, the imaging modality that might give a spurious result. If you get beyond that, it is possible that a change could be induced by the drug that could appear as a beneficial effect on the imaging modality but in fact, it might be entirely irrelevant. An example might be if we're looking at total brain atrophy and if the drug increased brain water, it might be possible that it would appear as if there is less atrophy when in fact, the change that was induced was entirely irrelevant.

The other possibility is that a drug may actually have an effect on a structure that might be relevant or that one might think would be relevant in the important pathology, but in fact that that effect might not be what we would think it was. For example, one could show that there might be less atrophy because, in fact, the treatment preserves neurons. But in fact that the neurons are not functioning properly, the beneficial effect on the picture might in fact be spurious with regard to its clinical concomitant.

So, with as a very brief background into sort of the regulatory framework in which we need to work and some of the conditions, I just want to pose to you the two large questions that we'd really like to discuss and ultimately vote on.

The first question is whether or not you think any of the imaging modalities that we're going to hear about this morning are in fact, or have in fact been validated in the sense that I've discussed and in the sense that Dr. Hughes will elaborate on. Failing that, we would like to know whether or not you

think it's appropriate for us at this time to rely on a drug's effect on an unvalidated surrogate to support the approval of an application for a treatment for patients with Alzheimer's Disease.

So, with that charge, I'll turn the microphone back to Dr. Kawas.

CHAIRPERSON KAWAS: Thank you, Dr. Katz.

Well, both the charge and the number of modalities that we're going to be reviewing today, and most notably the number of speakers that we're going to be listening to today, require that we keep this meeting as much as possible on time, which is already not happening.

For the speakers, I would very much appreciate if you could keep to your time. There will be a timer up there to warn you shortly before you will be pulled off of the podium with a hooked cane. And with that, I'd like to introduce our first speaker, Dr. Charles De Carli, who is going to give the overview of imaging.

DR. De CARLI: With that impossible task, I'll already start by saying that I am not going to be

able to accomplish it. First off, I?ve got to figure out how this works. There we go.

I want to thank Dr. Mani for inviting me and the individuals of our Committee here.

To talk about an overview of imaging is beyond this 15 minute time period that I was allotted. Instead what I'd like to do, as most of this will be reviewed more specifically by my other speakers, I want to talk a little bit about something we don't talk about that much, and that is understanding what is normal in imaging. We tend to focus on diseases and compare them to specific subgroups, but I would like to talk, just for a few minutes, about population based imaging and defining what is normal.

As you all know, data suggests that there's a linear change in cognitive performance with age, particularly in the memory sphere. What becomes obvious, however, after a careful longitudinal study which was done by the Chicago group, is that, in fact, what we see are individual differences in trajectory of performance suggesting that in fact the aging process is not monotonic descending, if you'd like to

say, but in fact has quite a bit of variability.

The process of aging involves multiple factors that include both genetic and environmental factors, and including lifestyle factors that may either reduce neuronal number in pair brain structure function or enhance neuronal number. However, ultimately over time with these risk factors and this balance we can lead to the process of dementia or not. And it's in that regard that I think we have to understand this very complex interaction in the setting of what is normal aging.

And for this, I would like to use some data from the Framingham study that I'll talk about to assess to certain questions for what is normal aging based on some cross-sectional differences that we see and rates of change to ask how do earlier life factors effect risk for later life dementia. And then the important thing, which we're not going to discuss today but will come up as a derivative of these conversations, and that is if we have surrogate markers, will we begin to use these markers very early in life to think about primary prevention strategies

and how to identify these risk factors.

All this data is based from the Framingham Heart study and funded by both the NIA and NINDS. And it's a community-based population study. The original cohort was begun to study in 1950 and continue about 400 of them continue to be under observation. Their children began to be observed in 1971. And this included routine assessment of cardiovascular risk factors, but MRI and neuropsychology was added in 1999. And we had the opportunity to cross sectional analysis as well as a repeat analysis in a subset of these individuals.

The quantitative brain imaging was based on intensity-based mathematical modeling to define segmentation of brain matter and CSF, white matter hyperintensity. And then we did some lobar analysis and also evaluated stroke volume in these individuals.

The cross-sectional data is a little over 2200 individual whose mean age was 64 years. But, as you can see, it's age range across most of the adult lifespan, 38 to 97 years.

As expected in an older population, there

was a slight higher prevalence of the females in the cohort.

This is example of the cross-sectional data. Individual data points plotted in men in blue, women in yellow and a regression analysis, a multi-variant regression analysis that includes looking at gender, age and age gender interaction, including a squared term. And this is just from here forward are going to be the regression models themselves.

Just to give you a sense of what the effects are, so this is on total brain cerebral volume of the hemisphere where we show a strong age effect, about 47 percent of the variance is ascribed to age with very little age gender interaction.

A similar relationship can be seen with the temporal lobe volumes. That is, there's a nonlinear decline with aging and no obvious age gender interaction. However, this seems to be slightly different when we come to front lobe volumes. In there, it seems to be an accelerated brain loss among men. And as my wife tends to tell me that if you don't use it, you lose it. And that's significantly

different than it is among women.

Ventricular volume can be seem as sort of the inverse of brain aging, in that the CSF spaces are increased with age in a nonlinear fashion. Again, there's no gender interaction.

And white matter hyperintensity volumes, which are in part an aging phenomena but also may represent cerebrovascular disease, show the only gender interaction effect. Of course, there's a nonlinear increase with age, but in about the seventh decade of life, you begin to see a differentiation between men and women with women showing greater volumes of white matter hyperintensity then men at a given age. And this has been shown in other studies.

The aging process is not only associated with degeneration, but also the appearance of cerebrovascular disease, which is quite common among individuals as they age. And here's an example of silent cerebral infarcts among this cohort. And you can see a steady age increase in the prevalence being slightly greater in men than women. In the tenth decade of life we just didn't have enough numbers out

here to make this reliable, so you tend to see this little bit of drop here.

So in the first part of this talk, it becomes clear that age accounts for approximately 30 to 40 percent of the differences in brain volume. And that it appears that brain regions change differently with change. That the frontal lobes, for example, do not appear to atrophy quite as rapidly as other parts of the brain. And this may, in fact, reflect a fact that these individuals in this cohort were essentially healthy. It'd be interesting to look at these changes in those who do not successfully age.

Gender differences, at least in this cross-sectional study appear to be modest. But the important fact that I think is becoming more and more recognized when we look at the consequence of aging in cognitive impairment, is that cerebrovascular injury was quite common in this cohort.

Next, for the final part of this talk I'd like to turn to the longitudinal evaluation of a very small subgroup of these individuals, 151, who were divided into five different age groups based primarily

on the fact that we had limited data available from this small study, 38 to 59, and then 60 to 69, 70 to 79, 80 to 84 and 85 to 96. And these two were chosen because we tend to see dementia beginning much more rapidly at these two higher age ranges.

We also broke this cohort up into or identified within this cohort 23 individuals age 62 who were at higher risk for dementia based on family history data; that is because of this, the Framingham study design, we actually know their parents? outcome. That is they either passed away without Alzheimer's Disease after the age of 80 or had it before age 80.

In addition, they may have one or both alleles positive for ApoE 4 and be at high risk for cerebrovascular disease.

We identified 21 individuals at low risk for cerebrovascular disease or dementia or having the converse, their parents made it to age 80 without dementia. They had no ApoE 4 alleles and had less an average cerebrovascular risk.

Within this longitudinal study design, there are MRIs repeated at about 2 years apart. And

these MRIs were analyzed separately and blindly by different raters.

This is an example of some of the data that we're seeing in this very preliminary analysis. What we find is that if you look at total brain volume, that there is age related increases in the rate. So this is the percent difference per year annualized change in MRI. I think this increase is a little bit spurious because of the small numbers.

Similarly, ventricular rate volume appears to accelerate, the rate of change seems to be increasing as we get older.

And finally, white matter hyperintensity, again, as possible evidence for cerebrovascular disease is accelerated in rates of change as the individual ages.

Now, in this very preliminary look at this data, the low risk offspring are compared to the high risk offspring. Again, these are people at age 62. And what we see is that they are essentially when you look at the rate of change of significantly different from a population mean of zero, there's very little

change going on in the low risk offspring with the possible exception of the white matter hyperintensity volume. However, in the high risk offspring we see a significant difference from zero and larger than the low risk offspring in rate of temporal volume atrophy and rate of increase of total ventricular volume.

And this is actual data. Again, these are only two observations so it's linear, and I think more observations will help clarify this. But what you can see in this hodgepodge of data is that there appear to be differing trajectories. Again, individuals who appear to be declining very steeply in terms of total cerebral brain volume; similarly there are individuals with white matter hyperintensity volume who seem to be going up much more rapidly. And we're even recognizing in small groups of people this heterogeneity in trajectories associated with apparent normal aging.

So in summary from the longitudinal data, it appears that the rate of brain atrophy in white matter hyperintensity accretion increases with advancing age. And I think that this has to be taken into account when you start looking at comparative

groups of individuals when you're looking at differences between dementia and normal aging. I think that if your normal aged group is 80 years old, it's going to atrophy more rapidly than, say, a younger group. And that may bring into contrast the differences between the dementia process.

Most importantly, I think, is we're beginning to recognize that individuals establish different trajectories of aging. And I'd like to suggest, although this data is quite preliminary, that this may actually begin quite early in life. And I think this is an interesting observation that has impact to using MRI as a surrogate marker, or PET, for that matter, or imaging in general in the sense that if we identify these individual differences, then we can study why these individual differences occur and possibly again to explore ways to modify these individual differences. Again, with the assumption that these changes represent a pathological process or an unwanted process.

So in conclusion, with 1 minute and 3 seconds, the understanding of use of imaging methods

as a surrogate markers for disease must include a clear understanding of what is normal.

And I thank you for your attention.

CHAIRPERSON KAWAS: Thank you Dr. De Carli.

Now Dr. Michael Hughes is going to talk to us about surrogate endpoints as measures of efficacy, complexities and limitations.

DR. HUGHES: Thank you very much for the invitation to speak here.

I'm actually a stand-in for Tom Fleming who was going to give this talk. And I must acknowledge him, because I borrowed a few slides from him. What I'm going to talk about is some of the complexities and limitations in looking at surrogate endpoints in clinical trials.

First of all, I thought it would be useful just to mention a few key criteria for study endpoints in trials. Clearly they must be measurable and interpretable in the context of the disease.

They also need to be sensitive to the anticipated actions of the drugs that you're

interested in. So, for instance, if you're studying an analgesic in terminally ill patients, you might want to focus on pain relief and not survival.

And thirdly, in terms of the approval process, I think they should be clinical relevant.

So here's a few examples of the difference between surrogate endpoints which tend to measure biological activity, some measures which aren't necessarily directly relevant to an individual patient versus those which measure clinical efficacy which are more directly relevant to individual patients that are taking the drugs.

You've already seen this definition of a surrogate endpoint. I've broken it up into two. The first sentence really deals with the idea that a surrogate is a substitute for one of these clinically meaningful endpoints.

And then the second really gets to the heart of what we mean by a surrogate endpoint in terms of drug evaluation. So you want to be sure that the changes that are induced by a therapy on a surrogate will reflect or will reliability predict the changes

in the clinically meaningful endpoint. And that's the hardest thing to validate in the context of surrogate endpoints.

Russ already mentioned some of the issues to do with or some of the interests in why we might want to measure surrogate endpoints focused particularly on drug approval, accelerated approval and full approval. But it's useful also to bear in mind that surrogate endpoints are useful for understanding the basic ideas of how the disease works and how drugs works. And they're also really pivotal in terms of Phase II clinical trials in deciding what drugs to take forward for further development.

So here are some examples of potential surrogate endpoints that have been used and the corresponding true endpoints. And this, again, I think brings out the idea that the surrogates often measures lab measures or other signs and the true endpoints are very much clinical endpoints which are relevant to individual patients.

The thing that I'd really like to stress first of all is this idea of what's the difference

between a prognostic marker and a surrogate endpoint use for drug evaluation. So we can think about a prognostic marker being any variable that predicts the clinical outcome. As there's no concept in that definition of effects of the drug, so there's no mention of interventions. Whereas, a surrogate endpoint is really something where the effect of an intervention on the surrogate reliably predicts the effect of the intervention on the clinical outcome.

So we're bringing in the idea that interventions effect the surrogate and, hence, effect the clinical outcome. And it's really critical to appreciate that a correlate, in other words a prognostic marker, may not necessarily be a good surrogate endpoint for drug evaluation. And what I'd like to do in this talk is really indicate why that's so.

So here's a very simple schematic of a disease which acts on or produces an effect on the true clinical outcome. And it also separately effects the surrogate endpoint, but the surrogate isn't on the causal pathway between the disease and the true

clinical outcome. There's going to be a correlation between these two.

So if you have intervention which effects the surrogate endpoint, it can have an effect on that endpoint without effecting the true clinical outcome. So a very common example of this is where the surrogate is a measure of symptoms of the disease and you can treat the symptoms without effecting the underlying disease.

So it's essential, really, to understand whether you've got just a correlation or an association, or whether you're dealing with a causal pathway between the disease and the true endpoint which involves the surrogate. So that involves a lot of basic science, clinical research to get at the pathways of the disease and also the ways in which drugs work on those pathways.

And then also empirical evidence about the performance of the potential surrogate in practice.

Now having said that, even when there's an established model for a causal pathway, you may not have a good surrogate endpoint. So here's what you

might think of as the ideal surrogate endpoint. So the disease effects the true clinical outcome via the surrogate endpoint. So if you have an intervention which effects the surrogate, you think it will also effect the true clinical outcome.

So the key thing here is that all mechanisms of action of the intervention on the true endpoint are mediated through the surrogate. But even in this setting, the effect of the intervention on the true outcome could be underestimated if there's a lot of measurement error in the surrogate. That's not a varied situation, but it does arise sometimes.

The other extreme is also very common. You can get overestimation of the effect on the clinical outcome if the surrogate effect is not of sufficient size or duration. And the key issue here is whether the effect of the drug might be transient or whether it will be maintained long term among the patients taking the drug. So these problems can arise even when the effect on the surrogate is statistically significant.

Now in practice, surrogates fail for

multiple reasons. And here's an illustration of one situation where it may fail. You've got the intervention effecting the true outcome via the surrogate. You've also got important pathways by which the disease effects the true clinical outcome, which aren't effected by the intervention. So the value of this surrogate in this setting will depend upon the relative importance of these different pathways, the ones which are affected by the intervention versus the ones which aren't affected by the intervention.

Here's another situation where the intervention actually effects the pathway which doesn't involve the surrogate. And in this sort of circumstance, if you rely upon the surrogate endpoint for evaluating the drug, you'll miss the true value of that drug on the true clinical outcome.

And here's one example of a disease, CGD, where there's a high risk of serious infections, recurrent infections. A clinical trial was undertaken to evaluate one particular intervention, interferon gamma. And those the surrogate endpoints or potential

surrogate endpoints in this setting were superoxide production and the ability to kill bacteria, so things which are relatively easy to measure.

But there was enough uncertainty about the value of these surrogate endpoints that a large scale clinical outcome study was done where the recurrence serious infections was the key endpoint in the trial. And in this particular trial they found a very dramatic effect on the true clinical outcome, but essentially no effect on the biological markers.

So this would suggest either the markers were just not sensitive to the effects of the drug or there were other important causal pathways relating the disease to the true clinical outcome, which the intervention actually effected.

So a key thing to appreciate is that if regulatory approval is based upon these surrogates, then that's going to focus throughout the evaluation on those surrogates. And if these surrogates are poor, then there's a possibility that you will miss drugs which have important effects on the true clinical outcome.

In practice, things are much more complex. You'll have not only the potential effect of the intervention on the pathway mediated by the surrogate, you may have an effect on other pathways which aren't counted by the surrogate that you're measuring. And you may have direct effects of the intervention on the true clinical outcome.

So in this setting the value of the surrogate is potentially unpredictable, and here's an example from the cholesterol literature. Here are two clinical trials which both effected cholesterol levels in roughly the same magnitude, so about just under a ten percent reduction in cholesterol levels in both trials.

The true clinical outcome of interest was all-cause mortality. You can see in one trial there was essentially no difference. In the other trial, there was actually an adverse effect on all-cause mortality associated with the cholesterol lowering drug.

So, this is despite the fact that if you look at the cardiovascular-specific mortality, there

does seem to be effects of the active drugs on that subset of deaths. So this would indicate that there are indeed either direct effects of the active drug on mortality and/or competing causal pathways which are effected by the intervention in an unpredictable manner.

Another example of a failed surrogate, and this is the classic one that's often cited, is anti-arrhythmic drugs which were widely used or prescribed post-M.I. to prevent sudden death. When a clinical trial was actually undertaken to evaluate these drugs with respect to their effect on mortality, it was found these anti-arrhythmic drugs actually tripled the mortality rate relative to placebo. So although they had their intended effect on arrhythmias, the effect on the clinical endpoint of greater interest was clearly adverse.

I'd like to finish with another example from the cardiovascular literature. And this looks at blockages of the coronary artery leading to myocardial infarctions. And the idea here was to evaluate the TIMI flow, which is a measure of flow through the

artery. And I've categorized it simply as complete flow versus no or partial flow.

And here are the results from one particular trial, which compared streptokinase to TPA. And you can see that the TPA increased the proportion of patients that have complete flow and also decreased mortalities, so you might anticipate that it's a good surrogate.

Then another trial was done to evaluate RPA, a new drug, versus TPA. And you see a certain slightly smaller effect on the surrogate. So you might anticipate that there will be a beneficial effect on mortality. But when a large trial was done to evaluate the effect on mortality, we found essentially no difference.

So although the surrogate had been predictive of the clinical effect in one trial, when it was taken to a different drug comparison it failed.

So in terms of the benefits and risks of using surrogate endpoints, the benefits have already mentioned. Clearly, we can do smaller and shorter clinical trials that will make drugs available sooner.

The risks of the drugs that are approved will have unknown effects on significant patient- relevant clinical outcomes. And the approval focused on the effects on surrogates could mean that clinically effective drugs are missed if the causal pathways are not well understood.

And it's important to appreciate that ultimately drug approval based upon the effects on a surrogate involves an extrapolation of experience with existing drugs to untested new drugs. Non-extrapolation will almost certainly mean that there will be an increased risk that drugs that are licensed could have no minimum effects or even potentially adverse effects on the patient-relevant outcomes.

So minimizing this really requires a very thorough understanding of the causal pathways for disease effects on the trust clinical outcomes as well as a similar understanding about the intended and unintended effects of all of the interventions, not just past interventions but future ones as well on the surrogate and the clinical outcome.

And you need empirical evidence to support

the validity of the surrogate. And I'll talk a bit more about that later.

Thank you.

CHAIRPERSON KAWAS: Thank you, Dr. Hughes.

We now have time for some questions for Dr. Hughes or Dr. De Carli.

DR. WOLF: I would like to ask Dr. De Carli, your studies were devoted to functional -- I'm sorry, to anatomical information. Did you also do any function studies and cognitive studies in order to correlate to what extent age was the only variable, and you indicated in your presentation and I realize it was a very short presentation, but were there also function studies that you're requiring to follow those patients to have a better view of, not only how the anatomical information is changing, but how the function information is changing?

DR. De CARLI: Yes, that's a very good question. Thank you.

Yes, we did -- I think the ultimate functional test we did neuropsychology. And so with brain changes we do see a decrease in some memory

functions and cognitive functions not only across the age spectrum, but also changing within. So within the very brief period of time that they were observed there were relationships between brain volume and general performance in cognition.

So we are seeing that. And, unfortunately, as Dr. Katz is reminding us, we're only see it one direction. If your brain is shrinking, so your point is going down. And the thought is, of course, is that we want to modify underlying processes and see if those people with underlying risk factors--for example, cerebrovascular disease, which we know there are proven treatments for, may alter those changes.

Does that answer your question?

CHAIRPERSON KAWAS: Thank you.

Please?

DR. PROVENZALE: I have a question for Dr. De Carli. Hi. You nicely pointed out the differences in brain volume changes in young at risk and young non at risk individuals. It appears from what you've shown us that if we studied a population comprised of

-- if we didn't identify the individuals who were at risk and combined non at risk individuals with at risk individuals we might mask or miss the effect of a drug. So do you advocate as part of drug trials targeting specific populations in that manner?

DR. De CARLI: Well, I think that, particularly in settings where you have a drug that you're looking at for primary prevention, I think that would be the focus of your study. You would take high risk individuals in which you were looking for them to progress onto a particular endpoint, be that mild cognitive impairment in an aging cohort or dementia in a cohort with mild cognitive impairment already present. And so, yes.

But I want to caution you that that data is very preliminary. It does coincide with some of the PET data that will be discussed, I believe, today and does support some of that data. But I still would emphasize that the sum total of number of individual studied, if you combine all studies available, are less than a 100. And you might gather that I have a suspicion of anything under 1000.

CHAIRPERSON KAWAS: Dr. Temple?

DR. TEMPLE: This for Dr. Hughes.

Would you distinguish at all between what one might call anatomic surrogates and functional surrogates? One intuitively feels that if you really have a good view of the anatomy, if that seems better. But I just wondered if you had any comments on that.

And then the other question I had -- or it's really a comment -- that surrogates fail for two potential reasons which are fundamentally different. One is that you were wrong about the relationship and the other is that the drug did something bad in addition to whatever the good thing was. I've always thought encaidide and flecainide reflected the latter. They're obviously drug were obviously pro-arrhythmic, so whatever good they might have done was overwhelmed by the fact that they were lethal. I wondered if you wanted to comment on that.

DR. HUGHES: In terms of the first question that you raised, I don't think there's a fundamental difference between how you would approach the validation of surrogate depending upon whether the

type of measure that it involves.

I think what that may effect is the type of study that you do to understand the basic science, the clinical rationale for the surrogate. But in terms of the types of empirical evidence that you would collect to validate the surrogate, I don't think you should have any effect.

CHAIRPERSON KAWAS: Dr. Van Belle?

DR. VAN BELLE: One comment to Dr. Hughes. I tend to agree with your remarks, and I think you'll see that later on. I think the correlation with the clinical entity is a necessary condition, but not sufficient. And I think we'll get into the sufficiency arguments later on.

I have a question to Dr. De Carli in terms of the measurement error of the total brain volume. From your graph you had in your presentations, I tend to see regression towards the mean. In other words, that people with very high brain volume initially tended to decrease and those with very low brain volume tended to increase a little bit. Can you give me some idea as to what the measurement error is in

this particular case? Thank you.

DR. De CARLI: Yes. I think that will be discussed in detail by some of my other colleagues. But in this very simple separate analysis, and this wasn't very sophisticated, the inter-class correlation for different raters on repeated measures is less than one percent. But that's a substantial amount when you're talking about a brain volume that's 1200 cc's.

But in a population, I remind you. In a population these effects can be seen quite easily.

CHAIRPERSON KAWAS: Dr. Fogel?

DR. FOGEL: Yes. This question is for Dr. Hughes.

I wanted to find out in the framework of the surrogate, how is side effects, for lack of a better term, factored into all of this in terms of true clinical outcome? And what I mean by that is, say, you have an anti-hypertensive drug and the endpoint, if you will, is decrease in cardiovascular disease, yet this anti-hypertensive at the same time causes depression. Do you consider that a successful surrogate or a failed surrogate?

DR. HUGHES: I think the cholesterol-lowering examples I gave were a good example of that. If you looked at the cardiovascular-specific mortality, then you saw beneficial effects in both trials. When you looked at all-cause mortality, you see no effect or an adverse effect. So that suggests that there are adverse mechanisms of action.

My own feeling is that one big issue is what is the true clinical outcome in drug evaluation. And that may be particularly hard in Alzheimer's Disease to think about. And there may be adverse effects on that true clinical outcome that could be due to the intervention. And in that setting clearly any validation of the surrogate endpoint will capture the potential for adverse effects. But I think there is always the possibility that there will be significant adverse effects which aren't captured within the true clinical outcome. And I think their standard approaches is for evaluating those adverse effects needs to go in parallel with approaches for evaluating the effects of the intervention of the surrogate endpoint.

DR. FOGEL: Well, I guess my question really is, though, that if the study is designed to decrease cardiovascular mortality, say, and instead the study finds that there was a decrease in cardiovascular mortality but there wasn't any change in all-cause mortality, how does one expect a surrogate endpoint to actually take into account all-cause mortality when, specifically, it's designed for a decrease in cardiovascular mortality. In other words, it seems that true clinical outcome in this framework is lumping everything. You know, how does the patient ultimately do in everything when the clinical trial is really just designed to decrease -- and I'm just using cardiovascular mortality as an example -- cardiovascular mortality and that's what the surrogate is basically being designed for, picked for, used for. But yet in this framework it's being considered as a failed surrogate even though it was not really designed to do all-cause?

DR. HUGHES: Well, I guess my opinion is that when you're trying to design a surrogate endpoint you should focus it on the very specific true clinical

outcome that you're interested in, so the disease specific outcome. And only if you're very uncertain about what the true clinical outcome should be should you consider broader classes of true clinical outcome.

So in the cardiovascular setting, if you're sufficiently uncertain about the potential mechanisms of action of interventions that you might be evaluating, then it might be important to understand how a surrogate fits in with a broader class of true clinical outcomes, including all-cause mortality. But I think the primary goal should really be to pick a surrogate which is validated in the context of the clinical outcomes, which are disease specific and leave the adverse effects of drugs as a separate issue which is routinely evaluated in clinical trials.

DR. FOGEL: So the cholesterol example then given that definition would have been a successful surrogate because it decreased cardiovascular but was neutral on everything else?

DR. HUGHES: Yes. And I think the reason those large trials were done was there was sufficient

uncertainty about how those drugs worked that all-cause mortality was thought to be a more appropriate outcome measure. And, in fact, all-cause mortality is simpler to measure and the effect to the cardiovascular mortality would be the very dominant component of all-cause mortality.

DR. FOGEL: Thank you.

DR. VAN BELLE: I have a question for Dr. Katz. We've defined surrogate, but we haven't really defined clinical outcome as to what is desirable.

Does the FDA have, kind of, a catalog of what constitutes clinical outcomes so that it can know what constitutes the clinical outcome? And the second part of that question is do surrogates sometimes become clinical outcomes?

DR. KATZ: Well, there is no catalog of clinical outcomes. But I think as generally understood, it's a measure of direct patient either functioning or subjective sensation of their symptoms, depending upon what the condition is that you're treating or some objective measure directly of how the patient is doing.

As someone pointed out, it's a measure that's relevant to the patient, him or herself, as opposed to a laboratory measure. Blood pressure is of no relevance to a patient in terms of how they feel or how they're functioning unless it's very, very low or very, very high.

And the second part was, do surrogates every become clinical outcomes? What are you thinking of specifically?

DR. VAN BELLE: Well, I was thinking on the context of, you know, blood pressure, where certainly action is taken to lower blood pressure and clinical action is taken just on the basis of blood pressure readings.

DR. KATZ: You mean in terms of practice? Well, sure obviously sometimes clinical interventions are employed entirely on the basis of laboratory -- drugs are stopped because somebody's liver functions are elevated. So in that sense, I suppose. But for the purposes of a clinical trial, clinical outcomes I would say are what we typically would use to assess the drug's effect. Clinical outcomes of the sort I

defined earlier.

CHAIRPERSON KAWAS: Dr. Temple?

DR. TEMPLE: But the surrogate never becomes the clinical outcome. We use it freely and comfortably, but it's never the clinical outcome for, if no other reason, that you can never know that the drug doesn't have some unpleasant side effect that you were not able to anticipate.

So if you do a blood pressure trial on 200 people, that doesn't tell you about a risk of one in a 1,000 of something nasty. So it can never be perfect. We use them, because we don't think those effects are likely.

I guess I want to make one other observation that clinical benefit is not free of the same risks. One of the great examples sometimes given of surrogate failures is the failure of several classes of drugs to treat heart failure to predict favorable outcome. In fact, none of those drugs were considered useful potentially because of their surrogate effects. They all improved symptoms of heart failure, but they were nonetheless lethal because they

did something else. And the same problem exists whenever you use a surrogate; you just have to collect enough data to reassure yourself on that point. But the surrogate never measures the other things the drug might do. It can't. Because you're not looking at those.

CHAIRPERSON KAWAS: Thank you.

Dr. Wolinsky and Dr. Sorensen, the last two questions.

DR. WOLINSKY: So I don't know if this should be directed to Dr. Hughes or Dr. Temple or Dr. Katz. But if the surrogate is basically to help facilitate getting answers more quickly that should be predictive of the clinical outcome, how do we feel comfortable about the safety issues if the trials are shortened by the effects on the surrogate?

DR. KATZ: Well, that's a real question. We're always -- for drugs to be given chronic --

DR. WOLINSKY: I wanted a real answer.

DR. KATZ: That, too, was a fair question.

Well, certainly for drugs to be giving chronically we are interested in knowing what the long

term effects are, the adverse effects of chronic treatment. And typically we would want that. Now, again, it's possible depending upon the nature of the treatment or the proposed indication and the importance of it, it's possible a drug could be approved without very much long term adverse event data.

On the other hand, adverse event data can be obtained -- often is obtained in the long term in uncontrolled settings, which is easier and sort of quicker to do. So there are ways to get that data.

Your point is well taken that if we have to do long term studies to get safety data, why wouldn't we want to do the long term effectiveness studies as well and look at the actual clinical outcome. But there are mechanisms if you really thought it was important enough to get this drug out there right away, we could have a minimal amount of long term effectiveness data with, perhaps, requirements in Phase IV to get more. And, again, you can get them in settings that are less onerous than long term controlled trials.

DR. WOLINSKY: Do we have good examples of rigorous Phase IV requirements for safety, as opposed to appropriate recognition by clinicians in the field of events that seem to be occurring at too high a frequency and that type of surveillance?

CHAIRPERSON KAWAS: Dr. Temple?

DR. TEMPLE: Well, sure, there are some wonderful examples. Cholesterol-lowering drugs, as you just saw, the early ones had great difficulty showing any benefit, perhaps because of a fluke or perhaps because they did something bad. But the statens, all of which were approved on the basis of lowering of cholesterol have, at least in four out of six cases, shown unequivocal major benefit in long term studies without much evidence of a downside that was sufficient to outweigh that. So those are really unequivocal.

There's also a massive amount of data on various blood pressure drugs.

I just want to observe that sometimes you can get your safety information from other settings. So, for example as people have pointed out, there

wasn't, maybe until recently, any study that showed that ACE inhibitors were actually good for you when you took them to lower your blood pressure. On the other hand, there's probably a couple of hundred thousand people randomized to trials in heart failure and other conditions. And in all of those there didn't seem to be any harm. So that might reassure you that when you use them to treat blood pressure, you shouldn't worry too much.

So a lot depends on what else you know about the drug from, perhaps, other sources.

DR. WOLINSKY: I think the question was actually a little bit more pointed, and that is, I think the studies that you've mentioned, which are very remarkable studies, were not necessarily Phase IV requirements.

DR. TEMPLE: They were Phase IV agreements in many cases, and it was before there could be requirements. And there's some question about whether outside of accelerated approval you can have requirements. But often people have been interested in doing it. So the big cholesterol studies were done by

companies that had agreed to try to do them.

DR. KATZ: There are also examples of drugs which have been approved on the base of relatively small safety samples, but for which there are registries in place in which post-marketing through which or in which additional safety information can be gotten for periods of time. So there are examples.

CHAIRPERSON KAWAS: Dr. Sorensen, and then we'll move on.

DR. SORENSEN: Yes, I have a question for Dr. Hughes. Dr. Hughes, I think you made the point that if you are looking at unknown, i.e, a surrogate, there's more risk than a known and presumably there's some potential opportunity for benefit or the regulations wouldn't have been modified to allow for that.

My question for you is are there any biostatistical tools to put error bars around the sizes of those risks? In other words, is there some way we can get a handle as we listen to these presentations about how good these surrogates are and

to get a sense for the goodness or the badness of that and what impact that may have on potential benefits or lack of benefit if we were to use that surrogate?

DR. HUGHES: I guess I'm going to talk about later about validating surrogates, but one of the things that you can get out of those validation procedures is some concept of how reliable the surrogate is, at least in the previous studies that have been done. I think the big unknown is whether the future study, future intervention fits in well with the previous interventions that you've studied.

DR. SORENSEN: It seemed like in the briefing material there some discussion about statistical tools to actually adjust for the difference between the surrogates you're using and the known outcome. And I don't want a biostatistical lecture, I'm just trying to figure out if there are some accepted tools that you might guide us through, maybe in your talk about that.

DR. HUGHES: Well, certainly you can estimate various measures which capture surrogacy and you can put confidence intervals on those and you can

use those as a guide to how reliable the surrogate is.

CHAIRPERSON KAWAS: Thank you.

We now have a block of speakers who will be talking about volumetric MRI and related subjects. And for the first one, Dr. Clifford Jack.

DR. JACK: First, thank you for inviting me to speak. I'm going to talk about structural MRI as a biomarker of disease progression.

And I'd like to begin by returning to a point that was raised initially by Dr. Katz, and that's that it's straightforward, but it's important to keep in mind the distinction between validating a marker of therapeutic efficacy versus validating a marker of disease progression. In the absence of a positive disease-modifying therapeutic trial, I don't think anyone can come up with evidence validating the efficacy of an imaging marker of therapeutic efficacy. On the other hand, we can muster evidence for imaging markers as measures of disease progression. And most of us in the imaging world, I think, are comfortable with the idea that indirect measures of disease progression can be validated, provided that there is a

plausible biologic link between change in the marker and progression of the disease itself and, if enough empirical studies, independent studies are provided that produce a common result, i.e., the measured tracks with disease progression.

And what I'm going to do here in the next 15 minutes is to present four different studies that I do think provide supporting evidence MRI markers as reasonable markers of disease progression, the first of which was published a number of years now. The objectives of this study were very simple, and that was to measure the annualized rates of volume change of the hippocampus and temporal horn from serial MRI studies in cognitively normal elderly subjects and people with Alzheimer's Disease and then to test the hypothesis that the rates were different.

Here are the two structures that we measured. The hippocampus and the temporal horn. The end was small in this initial study, but patients and controls were individually matched on age, sex and education, so those variables should not confound the results.

And here were the results. The annualized rate of atrophy of the hippocampus was 1.6 percent per year in normal controls, and in cases it was greater than twice that. The annualized rate of the expansion of the temporal horn was 6.2 percent in controls and in cases more than twice that. These numbers are negative, reflecting the shrinkage of the brain; these numbers are positive representing expansion of the brain -- expansion of the CSF spaces.

So our conclusion at this point was that this was a reasonable first step in that we did observe the expected differences in rates between patients and controls, but it didn't prove that, at a more fundamental level, that changes in imaging tracked or matched changes in clinical status in these patients.

And that was the topic of this next study, which was to test the hypothesis that a change on imaging, i.e., in this case rates of change of hippocampal atrophy over time from serial MRI matched clinical change. And we used the clinical transition or lack thereof as the gold standard measure of

clinical progression.

And I realize now that I should have put a slide in here describing mild cognitive impairment. For those of you who aren't familiar with the concept of MCI or mild cognitive impairment, it is an intermediate stage between cognitive normality and Alzheimer's Disease and most are all patients who eventually develop Alzheimer's Disease will go through a phase of mild cognitive impairment nearly always in memory alone or memory-isolated type impairment. And we can use, and others have used, this transition type analysis from normality to the category of mild cognitive impairment and on to Alzheimer's Disease as clinical measures of disease progression. It eliminates the reliance on a signal cognitive measure which, as I'll show later on, those can go up or down.

In this study we recruited a 129 subjects from our ADRC and ADPR grants which met criteria at baseline for either normal controls, mild cognitive impairment or Alzheimer's Disease. The controls and MCI patients could either remain cognitive stable or could decline. And this creates five clinical groups:

individuals who are normal at baseline and who remain stable; individuals who are normal at baseline but then who decline to MCI or AD; MCI patients at baseline who are stable or those who decline to AD.

And one can see that the age and the MMSE scores for each of these key parallelized comparisons, normal stable versus normal decliner, were equivalent. Same thing for MCI stable, MCI decliner; age at baseline and MMSE score were equivalent. So, again, these should not serve as confounding variables.

And here were the results. One can see that the annualized rate of hippocampal atrophy of normals who declined to either MCI or AD was substantially greater than that of normals who remained stable. The annualized rate of atrophy for MCIs who declined was substantially greater than that of MCIs who remained stable. And this rate was very similar to patients who started out with Alzheimer's Disease.

So the conclusion from this study was that the rates of hippocampal atrophy did indeed match the change in cognitive status over time. And we took this

as some measure of validation of the change of MRI volume as a marker of disease progression.

A next question, one might ask, is what about different techniques or different rate -- different brain measures. The question addressed in this study was, are some techniques better measures than others of disease progression and is there stage specificity. So the objective of this study then was to compare the annualized rates of atrophy by technique, and I'll describe different techniques, among six different groups this time: normal-stable, normal-converter, MCI-stable, MCI-converter and then AD-slow progressor, versus AD-fast progressor. This is defined on the annualized rate of change and the Mini-Mental score.

We measured four structures: hippocampus, entorhinal cortex, whole brain and ventricle. I'll skip over these for the sake of time.

Because this data -- we were warned not to show anything that hasn't been published yet, so some of the numbers here have been blanked out in this slide. But, what can you do.

If you look at these comparisons here, normal stable versus normal converter, and these are the four different measures of interest: whole brain, ventricle, hippocampus and entorhinal cortex. You can see that among normal converters the rate of change, annualized rates of change, are greater than those in the normal-stables for each one of these four measures.

Come down to this parallelized comparison, MCI-stable versus MCI-converter; again the annualized rates of atrophy for the converters are greater for each of the measures and then AD-slow progressor versus AD-fast progressor, these same results.

A reasonable question to ask then is do some of these measures perform better than others and is there some stage-specific sensitivity. And to address this question we use this metric, which is the difference in the mean rates between this group versus this group, for example, divided by the pool of variance. And if we then look at these four different parallelized comparison in rates with respect to the different measures, one can see that these three do

perform better than this one for this stage, i.e., normal-stable versus normal-converter distinction.

For MCI, again there seems to be a clear winner, and that's the hippocampal measurement. For AD-slow progressor versus fast-progressor and normal-stable versus AD-fast progressor this measure seems to be the best performer.

So from this we conclude again that structural MRI rates do seem to consistently follow expected correlations with clinical transition. And there does appear to be some stage specificity or difference in the sensitivity of these different measurements at different stages of the disease.

The last study I'll describe was a multi-site study. The first three studies I described were all derived from a single site. Any sort of a clinical trial, however, will be run via a multi-site approach. And so one can reasonably ask the question, can you get data that makes sense from multi-sites. And the data that I'll describe here was based on the Milamilene study.

The Milamilene study was originally

designed as a 52 week controlled trial of the this muscarinic receptor agonist. The therapeutic arm of the trial itself, however, was not completed due to a projected lack of efficacy on on interim analysis, but the MRI arm of the study was allowed to continue.

A total of 192 subjects from 38 different centers then ultimately underwent two different MRI studies separated by one year and we measured hippocampal and temporal horn volume rates.

This kind of study generates a lot of data, and I will only show 2 slides of actual data. These are the actual change data in five different measures. The ADAS-Cog was the primary outcome measure. MMSE and GDS were clinical/behavior ancillary measures. And then these are the imaging measures that were also used as ancillary measures in this study.

These are the annual raw change for each of these measures in their appropriate units, annual percent change and then this column is, perhaps, the one that's the most interest. So this is the proportion of the group that wound up declining on the

measure.

Now, again, all these people had mild to moderate Alzheimer's Disease, and so theoretically everyone of them should have declined because the disease was indeed progressing over this year period in every individual in the study. But you can see that the proportion of individuals who actually declined on the measure, it was only about two-thirds of the subjects on these measures, particularly this one which is the measure that's used in Alzheimer's trial.

Contrast that then with the imaging measures where decline was much more consistently seen, particularly with the hippocampus where essentially all people declined. An improvement in performance theoretically represents or can only represent an error in the measure.

If one does power calculations based on a 50 percent effect size, i.e., a 50 percent rate reduction over one year using these data, these are the data that we got. You can see that the estimated sample size requirements are substantially greater for

the cognitive measures than they are for the imaging measures. And this is entirely due to the much greater variance in the clinical/behavioral measures versus the imaging measures.

So from this study we then concluded that the technical feasibility of doing a multi-site trial with structural MRI atrophy rate measures was documented. It was validated.

The decline over time was much more consistently seen with imaging than with behavioral measures. And finally, due to much greater variance in rates for behavioral measures versus imaging measures, the sample sizes required were substantially greater for the behavioral cognitive measures.

This is the last slide I'll show, then, just to conclude by returning to the original comment that I made. And that is that in the absence of a positive therapeutic trial, a true disease-modifying trial that incorporated imaging, the best available evidence that we can muster supporting the validity of MRI as a biomarker of progression is multiple natural history studies that consistently demonstrate

concordant MRI and clinical change.

I'd like to acknowledge the Aging Institute for ongoing support of our program at Mayo, the members of our ADRC and ADPR grants, particularly Ron Petersen, my long time colleague and collaborator who was the principal investigator of both of these grants. These three individuals of Parke-Davis that allowed the MRI portion of the Milamilene study to continue, even after the therapeutic trial itself was stopped. And these individuals from my own laboratory.

Thank you.

CHAIRPERSON KAWAS: Dr. Jack, just for clarification, could I ask a quick question? What's the definition of decline on the ADAS-Cog? If two-thirds declined, do you mean two-thirds declined one point or more or two-thirds declined 4 points or more?

DR. JACK: Any decline.

CHAIRPERSON KAWAS: One point was enough?

DR. JACK: Yes.

CHAIRPERSON KAWAS: So only two out of three people in that trial even declined one point on the ADAS-Cog in 52 weeks?

DR. JACK: One or more. So that statistic just represents a positive change.

CHAIRPERSON KAWAS: Thanks.

Our next speaker on the series is Dr. Cecil Charles. Oh, I'm sorry. Dr. Nick Fox. My apologies.

DR. FOX: Thank you very much. I think it's a great honor to be invited over here, and I do apologize if anyone has any trouble understanding my accent.

I'm talking about rates of atrophy and in particular and try and follow on from some of the other speakers.

Just to give an overview of what I'm going to talk about, I'm going to talk about -- address the relationship of atrophy rates, the pathological and clinical progression in treated patients; it's the natural history points that Clifford Jack just made. I'm going to address the issue of disease modification versus symptomatic effect and then try and move on to the crucial question of whether or not it's reasonably likely that atrophy rate change would project clinical

benefit in treated patients. And that is very much related to this final point, is that the possibility that atrophy rates might be uncoupled from clinical benefit and how one might protect against that possibility if you might do so.

Okay. Just going back to the pathology for a moment, Alzheimer's Disease is characterized by the accumulation of tangles, plaques, synapse loss, dendritic pruning and cell loss and atrophy. And that is an inexorable, inevitable, characteristic, defining feature of the disease. And the relationship between atrophy and cell loss has been documented by several people, not my own work, where one has looked at a loss of hippocampal neurons and regional atrophy in the hippocampus in other studies.

Now, what can volumetric MRI address of these, I would argue, core components of pathology? Well, it can look at atrophy rates, and I'll talk about that.

First of all, what can we understand about the progression of the disease? Well, what I show here in this first panel is an individual who has just

presented with Alzheimer's Disease, clinical probable AD, so mildly effected. Then there's a further MRI scan 18 months later, and then a third another 18 months later. So a three year interval here. And what I'd like to point out, which is really just a pictorial description of Clifford Jack has just shown, which is this devastating loss of volume within the hippocampus here. But I'd also like to point out that what we see, if you look at the ventricular enlargement here, the sulcal enlargement, the Sylvian fissures here, that the disease is a region-specific progression from entorhinal cortexes on to hippocampus and on to new cortical areas but this process is well established even when people present to us in the clinic.

The technique which some of the results I'm now going to show, I'm not going to go into the methodological details it relies upon. Registration, that is positional matching taking a first brain scan, and super imposing a second brain scan very precisely upon it so you in effect fuse the two sets of data in the same spacial framework. You can then automatically

subtract those images, produce different images and create a direct measure of change in volume from those different images.

And just to show that descriptively here, the first scan here, somebody with mild cognitive impairment. And then what you see there is the progression over one year. And just go back for a moment, you can see the ventricular enlargement, the hippocampal loss.

Now, in Alzheimer's Disease, again, we can actually visualize the change. This is addressing the point about whether or not there is a signal there, is there something that we could measure. I think the answer is yes both in a region and a global way.

What does this translate to? This is, again, these are old published data. Now this is looking at early onset Alzheimer's Disease showing the rate of whole brain atrophy in age-matched controls with early onset. Alzheimer's Disease showing a very significant difference that's associated with the disease in terms of rate of brain volume loss.

Now, I've just struck these slides in

while I was waiting, so I'm sorry they're not in your handout. But to address one of the questions that was raised about the precision of measurements, this is a test in reproducibility using this test. It's a real test in that it's scan/rescan. So individuals had a first scan, then a year later they had two scans on the same day. And we looked at what the rate of whole brain atrophy going from A to B was when compared to A to C, which would expect to be very similar.

From that you can see there's a good correlation, and one can get a measure of the error in that measurement. And we have unpublished data, which is why it can't be shown here, with larger numbers essentially showing the same thing; which is remember that the whole brain volume is about 1200 cc's, typically in these patients. This is a .1 to .2 percent error.

Does that correlate with clinical decline? Well, yes it does. Shown here, rate of brain loss against rate of many mental state examination scores.

Are these changes consistent over time? Well, this is a group that have been prospectively

followed in a naturalistic way. And this is an ongoing study with multiple short interval scans.

This is looking at some of the 6 month data showing 2.2 percent in the Alzheimer's group with a standard deviation of 1.4 About .5 plus minus .8 in the controls. That's over 6 months.

The same individuals at one year show a very similar rate, but the variance is coming down, addressing this issue of feasibility and measurement of clinical meaningful change.

Now, if we turn to the individual we can show that within the individual that the measurements are sensitive enough to track change within an individual and that the changes in normal aging are very different to that seen in Alzheimer's Disease. This is percentage of brain relative to the initial scan.

Does it predict clinical outcome? Well, one group that we looked at was with individuals with a family history of Alzheimer's Disease. So these are individuals at risk by virtue by either a known single gene mutation or strong family history.

This is the normal aging change, which fits with my colleague's data, which he showed a moment ago, which is that whether its physiological and measurement error, there's quite a spread in the changes with just two time points in terms of rate of atrophy. And this is all based upon two scans.

I'd like you to look at this middle column here. These are individuals column here. These are individuals who are risk, again who just had two scans, and then who had been followed for 3 or 4 years following the second of those scans. Those individuals who remained well over that time period had these rates of atrophy and those who became clinically effected in the follow-up period a significantly greater rate of atrophy.

Now, I'd like you to look at this lowest point here, which I will now show you in more detail what then happened to that individual. So this is her serial imaging from 1993 to 1997 registered. In red you'll see the progressive loss of brain tissue or signal on the scans. So the early rate of atrophy that you saw in that shot was related to this pre-

symptomatic period here, which then slowly accelerates. But I think what this shows it gives you some sort of measure of some of the errors within the measurement, but also the inexorable nature of the decline.

That is a regional specific effect, and I won't go into the details, but the bottom pictures here are based on a technique called fluid registration. And in green you'll see areas which have the highest rate of atrophy on a local basis. So this is that same individual showing the progression from pre-symptomatic change here effecting the hippocampus, becoming more profound in the hippocampus, and then becoming more widespread by the time they're clinically effected.

So, just to summarize what I said so far. I think that pathological evidence shows that atrophy progression untreated AD inexorable correlates with cell loss. I think MR- based measures are reliable and sensitive to change, at least at the clinically meaningful level.

Now to address sort of what the meat of

what we're discussing here. We can say that rates of cerebral atrophy both from the previous speakers and my data, I think, are increased in Alzheimer's Disease. They do predict conversion to Alzheimer's Disease either from MCI or from familial cases at risk. Those rates of atrophy correlate with clinical decline. And I'd like to suggest that it's biologically plausible that regional specific atrophy reflects pathological and clinical progression.

Is this a disease-modification effect or a symptomatic benefit? Well, I think that the issues relating to staggered start or staggered withdrawal are similar. Is benefit sustained; are all disease effects modified? And one of the defining features of that is whether or not you're on the causal path, as it's been described, which is probably related to whether you're near the causal end of the process as well.

So I think that we suggest that these measures are feasible, they're sensitive, they're clinically meaningful arcing and changes seem to be correlated. But can change in one happen without

change in the other? The issue of a change in the surrogate is both a necessary and a sufficient condition for the clinical outcome.

Well, can one put forward a model that suggests that that seems reasonably likely? Well, it would seem to me the destruction of our neural networks is very closely related to cognitive decline and death. And that pathological process acts through that destruction.

But could volume change and thereby atrophy rates occur without that neuronal construction? Well, yes, it could. Because neurons are not the sole determinative of cerebral volume. Inflammation, hydration, osmotic effects, protein deposition can all change volume without changing the number or size of neurons.

For example, we've shown that hemodialysis can be associated with a three percent cerebral volume change over one day. So that's a worry. I would say, therefore, that neuronal changes are neither necessary nor sufficient to produce volume change. However, I think progressive volume loss is more likely, more

reasonably likely to be related to progression neuronal loss. So I would suggest one would require two or more imaging time points, perhaps many more, including if possible scans of treatment.

So I'd like to suggest that it's reasonably likely that a measure of slowed neuronal loss would predict clinical outcome. In fact, my whole model of how the pathology works in this disease is that if we could slow loss of brain cells, we would slow the clinical outcome, and that slowing would constitute disease modification.

I think it is reasonably likely that a slowed rate of neuronal loss would result in reduced atrophy rates or, the converse, that reduced atrophy rates in a properly constructed design would probably or reasonably reflect a slowed rate of neuronal loss. And if that reduction in atrophy rate followed both the region and time related pattern of pathology, so not just two scans which could be a purely a drug presence effect, but several scans in looking both at regional and global changes, then it would be reasonable to conclude that the clinical outcome is

likely to be improved.

So, that's just a summary. Atrophy rates correlate. Causality is plausible and it may be reasonable again with appropriate study design, to suggest this predictive power. However, disease modifying drugs are required to strengthen that link.

Thank you.

CHAIRPERSON KAWAS: Thank you, Dr. Fox.

Now I think our next speaker is Dr. Cecil Charles.

DR. CHARLES: This talk could be called the devil's in the details talk. I'm not going to show you any images, but talk a little bit about some of the issues if you're going to use imaging.

I'd like to also thank Dr. Mani for inviting me.

Basically, one of the things that you've already heard some information and you're going to hear other talks talking about imaging, and what we're talking about here is quantitative imaging. And the first thing I want to do is really kind of talk a little bit about what that means, particularly in the

context of quantitative imaging as opposed to what we most normally think of as clinical imaging on some of the issues related to the imaging protocols, issues of how you monitor these things in the multi-center trial and issues of analysis. And in the information that was sent out there was some information on how one might cross validate different analysis techniques.

There's sort of an interesting question that arises as we've tried to do this, is what is quantitative imaging and how is it different from what's done everyday by radiologists in the clinical field? Well, if you think about clinical imaging, most clinical imaging protocols are set up to visualize a disease, visualize a lesion or detect it, and it almost always has a radiologic interpretation. A radiologist looks at those images either on film or on a computer and has an output which is generally words to rule in or rule out this diagnosis. And certainly in this country they're going to have a primary, secondary, and tertiary diagnosis to make sure everything is covered.

What you've been hearing here then is

something a little bit different, and that's in quantitative imaging we're going to try to extract tissue characteristics of some kind from some imaging parameter, whether it's MRI, MRS, PET, whatever. And we're going to take that information and using some kind of algorithm, you've already seen some examples of how information goes from images back to numbers, and we're going to use this algorithm to basically extract numbers.

And the reason we want to extract numbers is because we want to be able to incorporate it in hypothesis testing. So it's the difference between looking at diagnosis and looking at effect monitoring.

Why does this matter? Well, it matters because the people that build these imaging devices, basically build them for this stuff over here. And this is something that's not really set up in the initial specifications or the design criteria for these kinds of systems. And it requires a little extra care if you're going to use them for quantitative imaging.

Is there a use for clinical imaging in

trials? Well, often times in many of these trials there will be imagings particularly for Alzheimer's Disease to rule out criteria, so there may be an inclusion or exclusion criteria. And that screening scan may not be done by a quantitative imaging protocol. So, subsequent imaging sessions are really going to have to be quantitative.

So what are the things that you can do with quantitative imaging? Well, you've heard about a lot of them, and one of the things that's nice about these techniques, especially magnetic resonance, is there are a lot of things that we can test. It's kind of the Swiss Army knife of imaging in that sense.

If we think about quantitative imaging, though, one of the first things we have to deal with is study protocol design. We're now trying to extract quantitative information so the way that we set up our scan protocol is going to be intrinsically different. Data quality, as has already been alluded to in these studies, Dr. Jack talked about the issues of doing this in multi-center trials, data quality is an issue. If the scan quality is not very good, then you're not

going to be able to analyze the data.

There's a lot of practical issues that I won't dwell on very long; data format issues, how data is cleaned rigorously and prospectively with criteria that are stated at the beginning of the study. Dr. Fox has already talked about data registration in serial studies. Fortunately we don't have to force people's heads into particular positions because the computer can take care of that after the fact. And there are many kinds of analytical protocols that can be used. You've already seen some examples of both tracing techniques, of boundary shift techniques, fluid mechanical registration, and there are a lot of them for extracting these quantitative information and ultimately the data's got to archived, which is certainly something that's becoming more interesting.

When we set out to look at these kinds of quantitative studies we try to work from the endpoint and say if we know how we're going to analyze the data, what can we do in defining the protocol to make it easier for the analysis algorithm, especially with magnetic resonance because we can change a lot of

parameters to change how the image appears and minimize certain kinds of artifacts.

We really also want to maximize this sort -- it should go without saying, but it doesn't always -- we want to maximize the amount of information content per unit time. We use in our lab a deterministic figure of merit where we simply look at the contract to noise ratio per unit resolution unit time when we're comparing different scan protocols. And the goal is to increase that figure of merit. And then you use that as, in fact, as a mechanism for testing the variance in your analysis algorithm. Because at the end of the day we can't change the effect size of a treatment, but we can work to decrease the variance both in the acquisition of the data and the analysis of the data.

And, of course, because we're working with human subjects, patient comfort and compliance is an issue. If they don't come back for the second scan, then it really doesn't work very well.

It's already been suggested that you can do this across multiple sites. Dr. Jack has certainly

done a study. If you're going to do this across sites, there's a lot of issues in imaging protocol cross-validation. You may work on different scan protocols, different scanning systems, you may work with different manufacturers. And even within manufacturers because there's different software and hardware platforms, you're going to have to rationalize the nomenclature so that when you're talking to a technologist running on brand S or brand G or brand P, and all the other brands, you're saying the right thing to the right person or they won't actually know what you're talking about.

Your site training needs to be uniform and uniform in the sense that, again, it has to be tailored to the particular manufacturer. Some people call a technique one name and other manufacturers call it a different name, even though it's the same physics.

Retrain with upgrades. In long trials people are continually upgrading these scanners and essentially watching this data as it's coming in in real time to make sure that the protocols are being

adhered to and that the scan quality is good.

Quality assessment's got to be quantitative, signal to noise ratio. Looking at issues like motion artifacts, which are very problematic in elderly patients, we use quantitative criteria, clutter to noise, to make that decision. And then these are cut off points for rejection based on the analysis algorithm. Some algorithms are more or less sensitive to artifacts. And that's something, again, you can define prospectively.

Protocol adherence, the obvious kinds of things.

And then system performance, and this gets back to the comment I made earlier that these things are designed to do clinical imaging and issues of spatial fidelity are a significant issue if you're going to measure the kind of brain volume changes that Dr. Fox and Dr. Jack have talked about, and you'll hear more from other speakers. If you're looking at clinical images and the field of view is 23 centimeters or 24 centimeters, it's not going to have a significant impact on a radiologist's interpretation

of diagnoses or detection of disease. If you're trying to measure hippocampal volume, on the other hand, it's a serious issue. So adherence to spatial fidelity is something that has to be looked at and it's not something that the manufacturers currently address at the level that we need in this kind of imaging.

So, incoming data formats. That's changing, it's getting better with DICOM formats, but there's still a lot of issue with varying media formats. You can't assume that everything's just going to be easily readable when it comes to your lab. There's all kinds of proprietary formats, and then there's also a lot of local formats, noncommercial PAC systems that have been developed. And these are more an irritation for anyone trying to deal with this data centrally, but it is one of the things that you have to deal with.

Data storage format, DICOM is really a data transmission standard. It's unfortunately becoming a file format standard and it looks like any file format standard that was developed in the '80s and ultimately, hopefully will improve it a bit. But

there are alternate standards that different labs use, and all kinds of standards that are out there. So there are many ways to store the data.

Data cleaning really has to be done with some prospective criteria of QC. You'd like to get rescans, if possible. If a patient moves and you get that data, if you can get them back in in a certain time window, because you want to minimize, again, lost data.

The data rejection has got to be on some quantitative basis, not just oh, this data looks bad. The clutter to noise exceeded a certain level, the signal to noise was not high enough and that's why you reject the data. And you notify the site and try to get the person back.

I'm not going to speak too much about data registration. Dr. Fox has already addressed it. You want to minimize positioning errors in the protocol. We always use immobilizers to minimize patient motion and to make it comfortable for the patient to hold still. Not trying to put them in any specific orientation, but letting them find a comfortable

position and use an immobilizer to try to keep their head there, and that tends to work very successfully. And as Dr. Fox pointed out, you can register serial scans in the computer. So with the computational algorithms that are there, you can minimize these problems.

And, again, even with on-site training, good technologists and so on there will be some misalignment and the computer can help us with that.

Analysis, well from the way we think about this, you need some perspective criteria of what you're looking at because the analysis is going to drive how you define your scan protocol. You want to optimize some type of figure merit, like I mentioned. You want to optimize your quality control criteria to match the needs of your algorithm. In some cases you may find that some algorithms are going to be more sensitive to these artifacts, and you need to address that.

And, of course, standard operating procedures. If there's any user interaction, if it's not a fully automated computational algorithm, you're

still going to have to deal with replicate analysis to address drift and look for an interrater variability. And you also will do this with computer algorithms, it's just easier there.

Archival is really going to be driven by the needs of the sponsoring agencies, and that's sort of an open question.

Central consolidation, replicate archival in different places, and at the coordinating center. An interesting question here is we're getting to very large data sets in some cases. And even though we think about things like CD ROMS and DVDs having very good reproducability, when you start talking about 5 or 6 terabytes of data, errors in one in 50,000 actually become pretty significant when you want to read that. So, duplicate, triplicate archival is going to be an interesting issue as we continue to look at these large data sets.

So basically in these kinds of things if there's a central coordinating center, it needs to have a close relationship with the sponsor, with the imaging sites also to make sure that the imaging sites

are really involved in the study. In many cases the imaging sites simply are seeing these people come through. They're not necessarily part of the treatment study and getting them sort of cognizant of why they're doing this so it's not just pushing the data through can be quite an issue.

Ongoing QC and quality assurance.

Blinded quantitative data analysis and with replicate analysis is critical. And to the extent that there are regulations that address this, of course regulatory compliance.

Thank you.

CHAIRPERSON KAWAS: Thank you.

And our final speaker for this sessions is Dr. Michael Grundman.

DR. GRUNDMAN: Thank you.

So I'd like to address a clinical trial that we're doing, the mild cognitive impairment trial that the Alzheimer's Disease Cooperative Study is doing. And we're using MRI as a potential surrogate marker in that study.

Mild cognitive impairment, as alluded to

earlier, is a transitional phase between normal aging and Alzheimer's Disease, and on a typical cognitive measure such as the MMSE, patients with a mild cognitive impairment generally perform at around a 27 to 28 range and decline at less than 1 point per year, compared to Alzheimer's patients who decline at 2 to 3 to 4 points per year.

Again, mild cognitive impairment, the clinical key criteria, in our study patients need to have a memory complaint which is verified by an informant with objective memory impairment. They generally have normal cognition other than memory and generally normal daily function.

Question: Why are we looking at mild cognitive impairment in our clinical trial? To begin with, part of the motivation for this is because we're trying to do a prevention study so we can delay the onset of Alzheimer's Disease. And previously we've shown in analyses from our clinical trial sites showing that patients who have MCI decline to AD at a much faster rate than normal controls, that way we can do this clinical trial with fewer subjects.

The clinical trial basically is designed as follows: We have 3 treatment arms, one with Vitamin E, one with Donepezil, one with placebo. We recruit the patients who have a memory impairment. And the goal of the study is to see whether or not one of these two agents can delay the clinical onset of Alzheimer's Disease over 3 years.

This is some of the details of the study. The doses of the agents involved. The study objectives, again, as I mentioned to prevent the development of Alzheimer's Disease, to slow decline on cognition and function and to see whether or not the agents might reduce the rate of atrophy on MRI. It's a 3 year trial, 769 participants with 69 centers in the US and Canada.

The baseline ADAS Cog in MCI subjects in our study is around 11 points. This compares to a normal control group that we're similarly following that has an ADAS Cog score of around 5. In other AD trials that we've done with our consortium where the ADAS Cog scores typically in the low 20s.

So, we're looking at several different

potential MRI outcome measures, but one that we've looked at so far or that we're trying to assess is the role of hippocampal atrophy since earlier studies have shown that it seems to be affected very early in AD pathology and contributes to the memory impairment.

So you've seen Cliff Jack's data before, which shows that patients who are normal controls have lower rates of hippocampal atrophy per annum than patients who are MCI, who are stable, who are decliners and then who have AD. And you can see that this might similarly parallel the decline that we saw earlier on the MMSE so that the rates of normal, the rates in MCI and the rates of AD are all somewhat different and in parallel to what you might see on a cognitive measure.

So the MRIs that we're doing in our study, we're doing them at baseline in a subset of the subjects, so there were 193. And then we're also looking at another scan at the time that the patients develop Alzheimer's Disease or they complete the study if they haven't developed Alzheimer's Disease. And we have a number of second scans, and this number is

increasing, but so far it represents only a subset. So I'm not going to discuss the follow-up scans yet.

So the specific neuroimaging hypotheses related to the hippocampus are that the hippocampal volume at baseline will correlate with the cognitive and functional measures, that it will predict who will develop AD, that the rate of volume loss may be greater in patients who decline clinically, and that the therapies might be useful in predicting which patients are going to decline.

So the first point from the baseline data which I think we've shown and published, is that the memory scores correlate with the hippocampal volume. This is just one example of the NYU Delayed Paragraph Recall Scores. And you can see that in patients with the smallest hippocampal volumes the scores on there, the number correct, were lower than the number correct in the patients at baseline who had the highest hippocampal volumes.

The other thing that seems to be apparent is that the hippocampal volume at baseline also seems to predict a conversion to AD and changing clinical

measures. The trial isn't completed yet, but the preliminary data that we have thus far suggests that people, if you just divided the group into hippocampal volume; the top half of the group versus the hippocampal volume of the lower half of the lower half of the group, you can see that the people in the top half of the group are declining, developing AD at a lower rate than the people who were in the bottom half of the hippocampal volume.

Similarly, you can see that on the clinical dementia rating sum of boxes, we have a similar effect where the people in the bottom half of hippocampal volume are increasing at a more rapid -- I shouldn't say at a more rapid rate, but they're -- over the course of two years they've reached a higher score, a worse clinical function than the people in the lower hippocampal volume.

So it was mentioned before what the optimal characteristics of a surrogate marker might be, specifically that the rate of hippocampal change correlate with the outcome measures, that it captures

its "net effect," and that it would be really nice if we could get the surrogate marker to show that it has an effect before the clinical decline or failure so that we could do shorter studies.

And ideally what we'd like to see is a close relationship between the rate of brain atrophy and the rate of clinical decline. And then we'd like a treatment that affects both of these things and is tightly linked to do that.

And so how might we actually do this in our trial? The plan would be to look at the slopes of decline in hippocampal volume or in whole brain measures over the period of the study. Take these slopes and then see whether or not the slopes show less rapid rate of decline in the nonconverters than in the converters. And then after we've done that, to see whether or not the treatment can also demonstrate a slower rate of decline than in the placebo group.

So brain atrophy, I think we've shown so far that the hippocampal volume is a good predictor of clinical outcome at this point, but it's possible that the brain atrophy may not always be a great surrogate

marker. For example, brain atrophy could occur due to weight loss or dehydration, and that could contribute to hippocampal atrophy. This was a clinical study that we did at our Alzheimer's Disease Research Center showing that body mass index correlates with mesial temporal cortex volume in both men and women. We published this several years ago. So this is just one indicator that there could be other factors besides potentially neuronal loss, for example body weight, that might also contribute to brain volume.

The other possibility is that the intervention may reduce the rate of brain atrophy and not improve the clinical outcome. So it could be, as Dr. Katz pointed out earlier, that agents that increase brain water or had an inflammatory response could lead to some type of brain swelling but not necessarily alter the clinical symptoms of the disease.

It's also possible that you could have an agent which reduced amyloid, and assuming that amyloid takes up some space, it could actually accelerate the rate of brain atrophy but at the same time reduce the

toxic effects of the brain amyloid. So at least in the short run it's conceivable, at least to me, that you might actually see an acceleration of the brain atrophy even though the drug might be doing something good.

What else?

So the other thing is that the intervention, in many cases you could have a symptomatic agent which could improve the clinical outcome, but not effect the rate of brain atrophy, And cholinesterase inhibitors are a good example of that. So if we relied only on a surrogate outcome measure, we might discard drugs that are good.

And then, of course, adverse events could occur that despite the fact that we see a beneficial effect on the surrogate which might ultimately make the drug ineffective.

And so I think hippocampal atrophy and brain atrophy both seem to be tightly correlated with clinical decline, but if we relied only on brain atrophy in the absence of clinical data, I think we should be cautious about that.

On the other hand, if we could show slowing of decline in addition to decline on clinical measures with a good safety profile, then slowing of brain atrophy might support disease modification claim.

And then finally, if we could do some clinical trials and we showed that the MRI data for specific agents correlated with both the clinical outcome and the rate on hippocampal atrophy, then it's possible that we could use that information in subsequent trials if we were to require two drug trials for two pivotal trials. For example, maybe we could accelerate that process and rely on the surrogate data in a subsequent trial and not necessarily require that all the clinical data be obtained.

And I'll stop there. Thank you.

CHAIRPERSON KAWAS: I want to thank all the speakers for their excellent presentations and for keeping to time.

The floor is now open for questions. Dr. Fogel?

DR. FOGEL: Yes. This question is for Dr. Jack. It?s actually a comment and a question. The four trials, the four studies that you mentioned, I guess in Dr. Hughes' framework would be more along the lines of a prognostic marker rather than a surrogate since there was no intervention?

And I guess the other question that I had was in one of the slides you said -- summarized in a number of the studies that the decline in imaging was more consistent than the behavioral cognitive measures. And the behavioral cognitive measures sounds like it would be the clinical outcome and the imaging is the surrogate or potential surrogate. So would that mean that that it wouldn't be in your opinion a good potential surrogate since it's out of proportion to the cognitive behavior?

DR. JACK: Good question. No, I wouldn't say -- out of proportion maybe is not the right phrase. What's really happening is that neurons are dying and synapses are evaporating in the brains of these patients. The question is what is the best measure of that pathologic process.

Is it a list learning, or you know a different cognitive test, or is it a measure of actual brain anatomy. In point of fact, in the studies you were alluding to, every one of those patients their brains were shrinking but yet some of them stayed the same or improved on these cognitive measures. That has to represent test error or retest variability. It can't possibly represent what's really going on pathologically in the brains of these patients.

The imaging measures simply, I wouldn't say, were out of proportion to the clinical measures. They were both trending in the same direction; downward, but they just did so more consistently across a large group of people.

DR. FOGEL: I'm not a neurologist, but couldn't you be having the plasticity of the brain and the neurons forming more efficient synapses and have the volume decreasing while their cognitive function either stays the same or improves?

DR. JACK: It's possible, but the way the disease works is the cells die, synapses die away and people's cognition in general goes down. I mean, I

suppose it's possible for synapses to regrow or whatever, but --

DR. FOGEL: Well, I mean as a pediatrician, we see a lot of kids who get totally devastated when they're born and we see their brains on MRI and it's incredible how much matter is lost. And yet you see them at 3, 4 years old and they could be nearly developmentally normal. I know that plasticity is different in children than they are in adults, but I also know that there are some plasticity that's been touted in adults as well.

DR. JACK: Agreed. I mean, you know, I think the issue of is there plasticity; if so, how much, how does it work and in adults is maybe not controversial, but poorly understood. Certainly there's much greater plasticity, no one would argue, in children than there is in adults. And, you know, I guess the key question here is to the degree to which there is plasticity, there is the capacity for brain repair in adults, can that rate keep up with the rate of pathologic progression?

The natural history of the disease would

suggest that in most cases it can't. In all cases it can't.

DR. FOGEL: Thank you.

CHAIRPERSON KAWAS: Dr. Katz?

DR. KATZ: Yes, I have a question for Dr. Fox.

You suggested that progression in atrophy is reasonably likely to be a reflection of progressive neuronal loss. But we are here concerned, obviously, with drug effects and what drug effects on what appears to be atrophy on an MRI might actually mean. So, would it be your view that a drug which induced a beneficial change in what appears to be atrophy could essentially only be due to an action of preserving normally functional neurons?

DR. FOX: No. The "only" is a crucial word in that question. Absolutely not. It could be due to a multiplicity of completely spurious causes.

But that's not -- only is a different level of proof to reasonably likely.

CHAIRPERSON KAWAS: Actually, I have a question that maybe is for all of the speakers, but

perhaps Dr. Fox could start out.

I think that we all pretty much are aware and agree with the data that hippocampal atrophy is associated with AD. We all agree that it seems to progress with the disease. But the piece of evidence that I'd love to hear from each speaker, they think is the best piece of evidence that suggests that stopping that atrophy will have a clinical effect on the patient.

It seemed to me when we got to that part of the discussion, then everyone started talking about reasonably likely and biologically plausible and seemed to me, and you know hair color and aging tracks. But dying one's hair doesn't necessarily do anything for aging.

What is the evidence that doing something to stop hippocampal atrophy will actually do something for the patient?

DR. FOX: I think since we've never managed to -- to my knowledge, since we've never managed to slow disease progression in Alzheimer's Disease, I don't think there's any way to provide, as

you say, evidence. Isn't that the only evidence that you will have?

I don't know whether somebody wants to --

DR. WEINER: I haven't been a speaker yet, but I would just add to that that there is quite a bit of data already correlating hippocampal volume with neuronal counts in the hippocampus. So I think it's fair to say that there's a lot of established evidence that as the hippocampus shrink it's because of neuronal loss. So if you had a treatment that slowed the rate of hippocampal shrinkage, one could infer that that was due to slowing of rate of neuronal loss, but it could be do to other things like --

CHAIRPERSON KAWAS: But it also infers, though, that the neurons are working.

DR. WEINER: Correct.

CHAIRPERSON KAWAS: Dr. Penn?

DR. PENN: I was very intrigued by what looked like a statistically significant effect on a number of your patients actually gaining hippocampal volume during the study, more than one would likely see just because of variation in the data. And this

brings me back to an example of a drug that's been shown to increase atrophy and then you withdraw it, and the brain comes back. And that's some work I did with Peter Carland in Canada 15 years ago with alcohol. It's very well established that if you drink a lot, your brain shrinks, and if you stop drinking your brain grows back, and the cognitive function follows those changes in brain size. That was done with CT and very primitive compared to the measurements you're now doing. But there are examples of a surrogate marker where we do have a change that very much goes along with the pathology.

So, what I'm wondering is are we obligated to look for those confounding variables in the patient population; that is nutritional status, alcohol, other things that might change hippocampal size when we're doing these smaller studies on patients with Alzheimer's Disease?

That's a question, sort of.

CHAIRPERSON KAWAS: Anyone in particular? Who do you want to direct the question to, Dr. Penn?

DR. PENN: Well, Dr. Fox, he looks like

he's nodding and off to sleep or something.

CHAIRPERSON KAWAS: Dr. Fox, you've got the floor.

DR. FOX: The nodding is just an unfortunate tremor, I'm sure.

I didn't show data on hippocampal volume change, but I do think that your example of alcohol is well taken. I think hemodialysis as I showed, if you changed the gases that people inhale, if you give people diuretics, all these things can change brain volume. And I think it is important that any study would look at other factors that might be confounders, but also I think it's very important that you look at the time course of the progression to try and see whether or not you've got a continuing effect on progression as opposed to a simple drug effect. I think that's important.

DR. JACK: Let me address two of those. First of all, in natural history trials the only way any of these alcohol or whatever are going to have an effect is if there is a bias in your study, and that is that if your control population has a different

rate of alcoholism or different rate of dialysis patients, or whatever, than your patient population.

With respect to a drug trial, I mean I think everyone here agrees that it is possible for a drug to dehydrate the brain or hydrate the brain, and that in turn will produce volume changes that are unrelated to any functional benefit.

The easy answer to the question, though, I think Nick was alluding this, is to at the end of the trial take people off drugs and determine if the drug produced a sustained change in brain volume that was not seen in the placebo group.

CHAIRPERSON KAWAS: Dr. Love?

DR. LOVE: Yes. Thank you.

This question is for Dr. Charles. You discussed a number of very practical approaches to an imaging protocol that you would recommend for inclusion. I would imagine you are speaking prospectively in developmental studies. Could you identify which ones of those you think are most critical in an imaging protocol going forward? And looking retrospectively if you were looking at the

literature, which key things would you look for in those articles to be sure the protocol wouldn't introduce too much noise?

DR. CHARLES: Well, all of them are important. Otherwise you're just adding variance. I mean, to the extent if you're looking retrospectively, I think you have to simply say -- and it'll show in the data. In other words, if the variance of the measure is higher, then that's likely due to combinations of issues of the analytical algorithm as well potential issues with site-to-site variance.

For volumetric measures within a single site that's well maintained, that works very well but you also have to check that over time because we all change our scanners over time. And we've seen in looking at our scanners at our institution that are well maintained, variations as much as 3 percent of the field of view in one-year time frames. And if you don't correct for those, again you don't have to fix them at the site, you just have to track them with a phantom so that you can correct the data after the fact and know what's going on.

DR. LOVE: Can I ask one related question to that? Also you mentioned that the software that's currently marketed may or may not be directly relevant. What approaches would you recommend to validate the software that's used across a multi-center study that's an add on?

DR. CHARLES: Well, the software is okay as it is up to a point. It's the combination of the software in the context that the goals of clinical imaging and the needs of clinical imaging are very different from what we're trying to do. So you have to add some additional materials like quality control phantoms that maybe the manufacturers don't provide. Particularly if you're going to go across manufacturer's boundaries, you can't easily compare a GE phantom to a Siemen's phantom, to a Picker to a Phillips, and all those other names so that no one will say I said the wrong guys.

But those kinds of things in spectroscopy, you'll hear more about MRS from other speakers, but there where you're dealing with very low signal levels, the way that we do studies is we actually run

a phantom at each setting. Because there's a lot of things that cause NMR signals to vary and you want to be able to track that over time. And you can remove that variance.

I mean, just as we do repeated measures designs to help minimize the impact of biologic variance, you've got to do something to deal with site-to-site and time variance with phantoms.

CHAIRPERSON KAWAS: Dr. Kim and then Dr. Sorensen.

DR. KIM: This question also relates to a little bit with what Dr. Love has alluded to. This question is going to be for Dr. Fox.

When you do your registration between the images, the pre and the post or the before and after, when you do correlation could you do some sort of normalizing before you actually measure your atrophy?

DR. FOX: If I understand you correctly, is the question could you deal with some of those scan adrift scaling changes or --

DR. KIM: Not just the drift, but there's atrophy overall. How do you take it out, what is

normal and what is disease?

DR. FOX: You can adjust for head size. For example, what is taken to be a rough measure of premorbid maximal brain size, namely inter-cranial volume. One certainly can adjust for that. As we all know that women are more intelligent than men, and yet they have an inter-cranial volume which is 12 percent smaller than men. So you can adjust for that. And when you do that, you can for example find that whole brain atrophy measures are appropriately accounted for and they match. You get rid of the gender difference by inter-cranial volume correction.

Is that the question that you were asking for?

DR. KIM: Yes. Because I'm looking for -- obviously most of us are looking for small changes. And I just wanted to make sure that those small changes doesn't get covered by the overall change.

DR. FOX: Well, I think most importantly is that the power of following the individual, the changes within the individual are what matter not changes between individuals. So what you have with a

serial study is the perfect control, i.e., the person themselves.

DR. KIM: Okay.

DR. SORENSEN: My question is for Drs. Jack, Fox and Grundman.

I mean, it seemed like Dr. Jack presented one slide that indicated that in three out of the four different groups of patients that he was looking at, that ventricular volume was more powerful or some way better than hippocampal volume. And I think I've seen that kind of data in the literature from other groups as well. And yet the other two speakers focused primarily, maybe not exclusively, but primarily on hippocampal volume. Is there a consensus among the AD imaging community as to, you know, sort of which single volumetric measure is the best one or is there a hope for that consensus? Or if there were going to be a primary outcome of a trial, would it be, you know, your five favorite volume measurements or is there a single one that we would pick, or how would you guide us there?

DR. JACK: That's an excellent question,

and the answer is that it's still unknown. And that is an active area of research.

I mean, the reason Michael showed hippocampal data is that -- for the trial that Michael showed the data for. The way the software worked is that you can measure the volume of the hippocampus or the inter-cranial cortex with a single data point. Our software algorithm is very much -- actually it was a knock off of Nick's boundary shift integral algorithm, and you need two different time points to put into the front end of the algorithm. So we won't have these other data, the whole brain regional volumes, et cetera, until both time points have been acquired.

But your point's right on the money; no one really knows. And it's very -- the point of the data that I was trying to show is that it's quite probable that the best measure will vary with the stage of disease. So early on in the disease one would suspect that measures of medial temporal load atrophy rates would be better, more sensitive to progression of the disease. Later on in the disease measures that were sensitive to atrophy in neocortical association

areas would in turn be better measures. I mean, that makes sense but I don't think anyone has really worked it all out yet.

DR. FOX: I've got very little to add to what Cliff said, except that what one's trading off is the amount of signal, which may change with the disease. So namely, if the most in absolute terms change was in an area, such as the hippocampus or entorhinal cortex at a particular stage of the disease, that has to be traded off with the measurement error or noise in that measure, which might be a physiological, it might be a measurement error of your technique.

And as Cliff said, the answer is not clear yet, but also I think speaking from a personal perspective, I would suggest if you're looking at probably a range of disease severities, because that's very difficult to characterize anyway, that one should be looking at at least a combination of measures. For example, a regional and a global measure. You can choose your region, you can choose your global.

CHAIRPERSON KAWAS: Dr. Katz?

DR. KATZ: Yes, I just wanted to follow-up to Dr. Fox's response to my question. Of course, he's right. I asked a loaded question when I used the word "only." But, of course, our standard is whether it's reasonably likely that an effect seen on a surrogate is going to predict the useful clinical outcome. And, of course, that's going to be a personal judgment, and everybody's going to make that judgment, I would imagine, on the basis of they would bring different information to bear.

So I would just second something that, Claudia, you said which is that if and when we get to discussing whether or not people think these things or other facts we might see on some of these measures are reasonably likely to predict the clinical outcome of interest, it would be very helpful for us to know what your individual personal bases were for deciding that something was or was not reasonably likely.

CHAIRPERSON KAWAS: Thank you. I want to start out then by rephrasing my question for all the speakers. What piece of evidence makes you feel that it is reasonably likely that the hippocampal volume

changes might be relevant to the outcome of the disease?

Dr. De Carli?

DR. De CARLI: Well, I think it's because we understand the disease process fairly well. If you accept the amyloid hypothesis that accretion of amyloid in the interstitial space leads to toxicity, neuronal injury and shrinkage, damage to neuronal trees and then subsequently loss of neuronal constituents that include axons; each of these phenomenon contain space. Okay? They're anatomically relevant structures that we can measure on MRI and that have high correlation. That MRI, the size of the hippocampus correlates with the anatomy, both increase in the pathology and loss of the tissue.

Now, it's --

CHAIRPERSON KAWAS: So to your mind what makes it most reasonably likely is just a strong correlation between volume and disease, is that correct?

DR. De CARLI: What makes it probable in my mind is that it's part of the pathological -- it's

the disease process. It's close to the end stage of the disease process that you have the -- the pathophysiology of Alzheimer's Disease is neuronal dysfunction followed by cell death. Now, structural imaging cannot measure neuronal dysfunction, but functional imaging may.

Second, however, is that that's followed by neuronal cell death which structural imaging measures quite well. So since it's part of the cascade, that if you stop or you interrupt cell death, then you therefore would stop the atrophy process.

Now, does that mean you couldn't have long term improvement symptomatically without these anatomical changes? Of course.

CHAIRPERSON KAWAS: Well, actually, the question's more the opposite.

DR. De CARLI: Right.

CHAIRPERSON KAWAS: Can you not have any real improvement in the person even if you stopped this change is really the concern? We all agree that symptomatic therapy shouldn't change by definition. But the question I think we're grappling with now is

if we stopped this atrophy with whatever compound in whatever way, how confident do we feel that we will see that reflected in the outcome?

DR. De CARLI: And my short answer is I feel very confident if it was involved in the cascade of pathology, and that's what I think the imaging is measuring. So if you have something that you know effects the cascade of pathology and you see this, then my confidence level would be extremely high.

CHAIRPERSON KAWAS: Does anyone else in the group want to --

DR. De CARLI: Anyone else want to stick their neck out? This is a public hearing, go ahead.

CHAIRPERSON KAWAS: Yes. Yes.

DR. FOX: I'd like to sort of try and give a sort of more considered or sort of detailed answer to Dr. -- not to your observation, but to my previous one, Charlie, which is Dr. Katz as well, which is your point is very well taken. And I think for a start one has to reasonably likely -- I mean, are we talking about -- are we all talking about 51 percent is one issue. I think maybe one should look at what

percentage one's looking at for the reasonable likelihood. Obviously, it'll be a guesstimate.

But I think that to any answer about that reasonable likelihood or what pieces of evidence, would have the caveats that I would want to see pertinent information about the design of the study. So a sustained effect. An effect that looks like it's both regional and global, an effect that has maintained beyond withdrawal of the drug and, as much as possible, of the confounders that are coped with or adjusted for.

CHAIRPERSON KAWAS: Thank you.

Dr. Temple?

DR. TEMPLE: I have what I guess is a practical question. One of the things that makes the a surrogate start to look really good is a successful drug. So people stopped worrying too much about blood pressure once the VA did its studies. Nobody remembers this anymore, but prior to those studies there was a huge debate about whether lowering blood pressure was good for you or bad for you. The so called New York School assured everyone that you'd

have more strokes and more heart attacks if you lowered blood pressure. That was not tenable anymore once the VA did its studies.

As a practical matter it's hard for me to imagine trials that would not include a certain number of clinical observations. I mean, we know you can show effects on cognitive functions in studies of modest size with drugs that work only a very little bit. So there's been success in there. They're not backbreaking trials, you can do them.

So at least early on my thought would be that people would be studying at least some patients who had observable disease and conceivably there'd be some interest in people who weren't sick yet. It's that latter group where it would be tempting, I suppose, to rely entirely on surrogate data. But wouldn't much of the support for doing that come from the fact that you'd been able to show something in people with already developed disease, which historically, at least, doesn't seem that hard. Maybe then after that additional claims or things like that might be based on the surrogate finding.

But are we really going to be looking at case where there are no clinical data? That seems unusual, except when you're trying to maybe stop people who aren't sick yet, that's the one case where you might take a very long time to get real data and might therefore want to rely on surrogate data only. But just as a practical matter won't some of the confirmation that the surrogate is plausible come from the observed clinical effects in the people who are already ill?

CHAIRPERSON KAWAS: Go for it.

DR. WEINER: I think you've -- that's the whole point. Nobody is talking about using imaging as a primary endpoint right now for these trials. At least, I've not heard any rational person say that we should right now start using imaging as a primary endpoint.

The role of imaging right now is going to be to provide confirmatory evidence to the primary endpoints of ADAS Cog and other confirmation.

So when the Phase III clinical trials are done, they will be powered for ADAS Cog and other

clinical endpoints. And I predict that imaging will be used as perhaps a subset of some of those subjects to provide, first, confirmatory data which is important in the regulatory process. And, secondly, the critical issue is whether or not the drugs have a disease modifying effect. And it's very difficult in the Alzheimer's area when you're using ADAS Cog to demonstrate disease modification with clinical measures alone. The only way to do it rigorously is to do a randomized withdrawal or a randomized trial which requires very large samples, takes a long time and costs the companies a great deal of money. There's a lot of dropouts in these kinds of studies.

So, if one designed a Phase III trial powered for ADAS Cog to demonstrate a clinical effect and used imaging to demonstrate disease modification, that's I think the role of imaging. That is, if you show that a drug slows the rate of cognitive decline, the same time you show that the drug slowed the rate of hippocampal volume loss, and finally if you show that there was a correlation between the ADAS Cog effect together with the effect on atrophy, that would

be I think fairly compelling evidence that your marker is providing evidence for disease modification. And I think that's the current role in Phase III trials.

CHAIRPERSON KAWAS: Dr. Katz?

DR. KATZ: Yes. I think this discussion is very good and very important, and certainly something we need to hear. I think we may be having it a little early in the day. I know that Dr. Hughes is going to talk again about validating surrogates; I think that's an important thing for people to hear, whether or not a single trial with a single drug that shows the correlation between clinical and imaging is sufficient to validate even that drug as having an effect on progression is an outstanding question, let alone for that marker for the field in general.

So, I hate to cut off discussions, but I think we might profit more from this after the speakers have been heard.

CHAIRPERSON KAWAS: Excellent. Thank you. Thank you very much.

And in fact on that note, how about a 15 minute break.

(Whereupon, at 10:46 a.m. a recess until 11:08 a.m.)

CHAIRPERSON KAWAS: Thank you, and we'll be restarting. And now we will be moving to a section on MR Spectroscopy and PET. And our first speaker is Dr. Michael Weiner.

DR. WEINER: Thank you very much.

I'm going to be talking about the use of MR spectroscopy and MRI to measure treatment of Alzheimer's Disease and neurodegeneration.

So what we need are imaging surrogates which are specific measures of neurodegeneration. We've been talking a lot about that this morning. And we also need sensitivity. We want to have maximum statistical power to determine treatment effects, fundamentally because the clinical measures have so much variability that huge numbers of patients are needed in order to determine treatment effects. MR spectroscopy, perfusion MRI and structural MRI are all candidates here.

Magnetic resonance, spectroscopy measures metabolites in the brain and a metabolite called N-

acetyl aspartate or NAA, which is located almost solely in neurons has been thought to be a measure of neuronal number or density, which would be a good measure of neurodegeneration, but it's also sensitive to changes of neuronal metabolism. And you'll hear more about that from other speakers.

Spectroscopy also measure colon metabolites, creatine myo-inositol which will also tell you something about what's going on in the brain.

We have been using a multi-slice magnetic resonance spectroscopic imaging technique illustrated here where we display images of that normal marker NAA, creatine and choline, and one gets spectra from individual regions of interest as shown for example here from white matter or gray matter. This large PQ represents N-acetyl aspartate, NAA.

An example of the kind of data you get from doing these studies, is this is a cross sectional study looking at about 40 patients in each group with healthy controls, Alzheimer's, subcortical ischemic vascular dementia, and patients with cognitive impairment. And what this slide shows is the NAA

concentration in the hippocampus and in the frontal lobe in these four groups.

Now, in the hippocampus you see the healthy controls have high levels of NAA. It's reduced quite substantially in Alzheimer's Disease. It's not reduced as much as ischemic vascular dementia. And on the cognitively impaired subjects, it's reduced about the same amount as the Alzheimer's patients.

On the other hand, if we look at the frontal lobe, note that the patients with subcortical ischemic vascular dementia have a much lower NAA compared even with the Alzheimer's patients. And this different pattern contrasts with what we see in the hippocampus where in Alzheimer's Disease the NAA is lower than in subcortical ischemic vascular dementia.

So you can use spectroscopy to get different patterns of metabolic change which characterize different diseases.

Now, in treatment trials, of course, we want to do longitudinal studies to determine treatment effects. And these are some data from a relatively small sample in our lab showing changes of NAA shown

here and choline shown here in controls cognitively impaired subjects and patients with Alzheimer's Disease. And what you see is the rate of change of NAA in the Alzheimer's patients in both the frontal and parietal cortex are greater than those in controls and that the cognitively impaired patients have an intermediate rate of change.

Interestingly, choline is also showing changes in the Alzheimer's patients similar to those seen with NAA.

A number of studies have been published using longitudinal MR spectroscopy, and you're going to hear more about this from the subsequent speakers, but the number of studies were small and I personally believe that currently we really have insufficient data concerning MRS as an outcome measure for longitudinal studies and Alzheimer's Disease. We just don't have enough data to say whether or not spectroscopy is going to be useful.

Now, another candidate is arterial spin labeled perfusion MRI. This is the technique that measures cerebral blood flow in the brain

quantitatively by magnetically labeling the blood that flows into the brain and then performing an image of the brain which detects the rate of blood flow. And these intervals of what these sort of images look like in healthy elderly controls in patients with Alzheimer's Disease, and this is some early data from our lab showing in elderly control subjects the rate of cerebral blood flow in the frontal, parietal, temporal and occipital lobes showing very substantial reductions of blood flow in Alzheimer's Disease giving you the magnitude of the decreases and the effect size. No one has done to our knowledge longitudinal studies of arterial spin labeling in Alzheimer's and we have no idea whether or not this is going to be a useful measure in clinical trials. But this gives us the kind of information you get from PET scanning. It only takes 12 minutes, so it's possible that this could provide that kind of data within the context of an MRI examine.

Structural MRI, we've been talking about it, it has phase validity as a measure of neurodegeneration. There are different measures of

brain atrophy. Nick Fox developed the boundary shift interval method and Cliff Jack has talked to you about the hippocampus. Data has been reported on different groups of subjects and it's been hard to compare methods, as we pointed out.

We've done a study on 23 elderly controls and 19 Alzheimer's patients who were studied with two scans and with a mean interval of about 2 years. And this gives the rate of change of the entorhinal cortex, the hippocampus, several different measures of the boundary shift interval, the cortical measure, the ventricular measure and the total brain atrophy measure. And this is a measure of the rate of change of the cortical gray matter measured by segmentation.

You can see that the controls have relatively low rates on the order of one percent per year. The Alzheimer's patients, depending on the measures, the entorhinal cortex, 7 percent per year, a very high rate of atrophy. The hippocampus about 6 percent per year. The whole brain measures have smaller rates of change.

And the coefficient of variation of the

Alzheimer's patients is shown here, and the statistical power to measure a treatment effect in Alzheimer's Disease roughly scales with the coefficient of variation. That is the variants in the Alzheimer's population.

This is the beginning of what we need to do for validation, that is this is the rate of atrophy of entorhinal cortex shown in aqua or hippocampus in white plotted against the Delayed List Recall score, which is a measure of memory. And basically what this shows is that patients with relatively good memory have low rates of atrophy. And the worse the memory, the higher the rate of atrophy.

This is some rough calculations of sample size for a 20 percent treatment effect in one year with different amounts of power, 80 percent power or 90 percent power or using a one tail or two tail statistic depending on your a priori hypothesis. And this would be sample size per arm.

So what this shows is that for the entorhinal cortex and hippocampus one can detect a 20 percent disease modifying effect with something on the

order of 50 to 80 subjects per arm in one year. That's a lot more power than you would have if you did this using ADAS Cog. You'd need maybe two or three or four times that number of subjects.

This is a way to analyze the structural imaging data using something called non-rigid transformations where you take two scans at time point one and time point two. And then using a computer program which essentially warps the scan from the second time point back to the first time point so that every individual pixel in the MRI is coregistered back to the first time point. And this shows a picture of the shape change that occurs between time point one and time point two; the blue showing contraction in the cortex of the brain and the yellow and red showing expansion of the ventricles and the CSF.

This shows how you could do that sort of warping between time point one and time point two in a whole series of subjects, and then warp these change maps to a common space so that one could essentially have a measure of change for a group of subjects. And this compares the changes in 55 cognitively normal

subjects with two scans versus 17 Alzheimer's patients with two scans.

Note that the scales are quite different because the Alzheimer's have so much greater rates of change. The main point to make is the Alzheimer's patients have much more change in the median temporal lobe region shown here, shown here and shown here than you see in the controls. So we do have a pattern of more rapid contraction in the median temporal lobe in Alzheimer's. And the beauty of this approach is it's completely automated. It looks at the whole brain and allows you to do both hypothesis testing as well as explore studies to look for regions of contraction.

Another way to display this same data is to look at a surface rendered image. This is a contraction in controls without lacunae, 37 subjects with an inner scan interval of about two years. The blue shows contraction in the cortex. Over here we're seeing 21 Alzheimer's patients with a lot of cortical contraction. And these are controls who have lacunar infarcts who are completely cognitively normal start showing some increase in the rate of contraction of

these subjects.

Another beauty of this kind of warping approach is that you can correlate cognitive change with shape change. So what this image shows is a correlation between the change of brain structure over time with the change of the mini-mental state examination over time in Alzheimer's patients showing those brain regions which had a significant correlation with the mini-mental state examination.

So in other words, it's kind of an image oriented approached towards, you could say, the beginning of surrogate validation here. Because we're correlating the surrogate, the image, with the primary measure of the cognition. And it shows that there are certain regions of the brain that are more correlated with the mini-mental state examination, and interestingly, more on one side.

So in conclusion, structural MRI has high power to detect longitudinal changes in Alzheimer's Disease. Structural MRI is a relatively specific measure of neurodegeneration because it's probably not very effected by brain activity or metabolism. In

contract, as you'll hear from the subsequent speakers, PET and spectroscopy are sensitive to measures of brain activity and metabolism; that's the power of spectroscopy and PET scanning. But that's also their disadvantage, that because they are sensitive to state they are less specific measures of neurodegeneration.

Structural MRI does correlate with cognition, as we've shown and Nick and Cliff Jack have shown, but much, much more work is needed to correlate structural MRI with cognition.

Certainly this is all useful in Phase II. It's currently an unvalidated surrogate. It's not a primary outcome measure for Phase III trials, but structural MRI is useful to provide confirmatory evidence using that FDA regulatory language of effect and to provide evidence of disease modification, which is what we need as this new class of drugs enters clinical trials.

What is needed are standards for MRI and spectroscopy and PET so studies can be compared. Because currently different investigators are all doing it different ways and it's really hard to

compare data. We need to have more correlations of imaging data with cognition function and pathology, and we need data from multiple sites for powering of future trials.

Cliff Jack showed the beginning of that with the Milamilene trial, but we need more of that.

So in order to get that, what we need is a longitudinal, multi-site observational nontreatment trial of controls MCI, NAD, using MRI and PET along with cognition and biomarkers. And, hopefully, a study like this is ultimately is going to be supported by the National Institute of Aging with co-funding from the pharmaceutical industry.

Thank you very much.

CHAIRPERSON KAWAS: Thank you.

Our next speaker is Dr. Murali Doraiswamy.

DR. DORAISWAMY: Thank you very much.

I want to thank Dr. Katz and Dr. Mani for inviting me here, as well as the advisory panel for inviting me.

I'm going to speak on MR spectroscopy. And many of the studies I'm going to be presenting are

relatively small sample size studies. And I want to put that in context. This is an issue that has not been discussed, which is the cost and the time it takes to do these MR studies.

A typical MR exam may take about an hour of the patient's time, perhaps a whole day of the experimenter's time to plan the protocol and to analyze and extract the data. They're also very expensive.

And a 10,000 patient clinical trial that has two MR scans, one at the beginning and one at the end, say 20,000 brains, is going to take a very long time to analyze. Because a typical academic lab processes about two to five scans a day if they're very efficient. So you can see if there are 20,000 scans, it's going to take a very, very long. So it's not the same as doing exams.

And some of the limitations in the longitudinal studies and the sample sizes we're seeing today are really a limitation of the expense and the time it takes to do these studies. And, hopefully, the NIA initiative would address that.

So as the previous speaker mentioned, brain MR spectroscopy is a non-invasive technique that provides a biochemical window into the brain and it can look at concentrations of metabolites either in whole brain or in discrete regions. And the size of the discrete region you want to look at is partly the limitation of the technique.

Now one of the important things to keep in mind is that MRS is usually acquired along with an anatomical MRI image. So really at perhaps at ten minutes or more you can get an MR spectroscopy scan in the same sitting that you get an MRI scan. So really you can get synergistic information.

Now, there are a number of MR spectroscopy markers depending on the type of MRS study that one undertakes. The type of MRS that I'm going to talk about is called proton MR spectroscopy or one hydrogen spectroscopy. And really the two markers that people are talking about with regards to Alzheimer's Disease is N-acetyl aspartate and Myo-inositol.

Now the key point to keep in mind here, again this goes to the heart of whether this

constitutes a surrogate marker or not, is we still don't understand fully the function of N-acetyl aspartate in the human brain. There is increasing evidence that it's an acetyl donor involved in various lipid metabolic pathways, perhaps involved in cell membrane, neuronal axonal membrane and in other kinds of neuronal functions, but we still don't understand it fully.

So without understanding the function, it's hard for me to stand up and say that it's truly involved in the causal pathway of Alzheimer's Disease, even though we don't even know all the causal pathways of Alzheimer's as well.

It's abundant in the human brain and some data suggests that it's the second most abundant amino acid in the brain. So common sense suggests that it is involved in a lot of fundamental processes. It increases during brain development.

There is a variety of postmortem and histochemical studies using specific antibodies that have shown that N-acetyl aspartate tends to be concentrated largely in the gray matter regions of the

brain. It's primarily present in neurons and not as much in glia cells. And this has also shown in culture, cell culture studies. So it's present in gray matter to a greater extent than it is in white matter or in CSF.

And that is really what goes to the heart of the postulate that it's a marker of neuronal function or density. And there's two kinds of studies. The earlier studies suggested that it might be a marker of neuronal density, and these were studies that correlated histopathological sort of changes and did postmortem MR spectroscopy, but there's more recent clinical evidence suggests that it may be more a dynamic functional marker rather than a marker of neuronal counts or density.

Now, the other marker that's of emerging interest is Myo-inositol. Again, we don't know exactly what this marker does or what it represents. There are many theories. Some people say it?s a constitute of cell membranes. But really there is recent evidence, at least suggesting perhaps it's a marker of glial activation. And there's some data

suggesting that it's increased in the prodromal stages of Alzheimer's, such as in patients with MCI or in patients with Down's Syndrome who haven't yet developed the Alzheimer's.

Now, you have to put these markers in perspective, and this may or may not be a popular slide, but I think it's a slide that everybody on the Committee needs to be aware of.

Now, the reduction in NAA is not specific for Alzheimer's Disease, as has been referred to by several people who have talked about really body weight, there's a number of other factors, but really there's a wide range of diseases effecting the brain in which NAA has been reported to be reduced. Now, I'm not saying that all these studies are very rigorous good studies. By and large, they're small. By and large, they're cross sectional studies. But a number of different conditions.

So, again, suggesting that NAA if it's involved at all in the pathophysiology of Alzheimer's Disease is more a downstream marker rather than something that's early and very, very specific for the

disease.

Now all the conditions that I have marked by an asterisk, including some that I've not indicated, are conditions where potentially reversible changes in NAA have been reported after either therapy or spontaneous recovery. And I'll give you one example.

In temporal lobe epilepsy, sometimes surgically they take out the effected seizure focus. And when you look at the contralateral side, NAA levels increase by about 50 percent after about 6 months after surgery and up to 100 percent a year after surgery in some studies.

Now, these are all the conditions in which hippocampal volume has been reported to be reduced. And one of the conditions that's very interesting, Cushing's disease characterized by high levels of cortisol, and there's very good animal data suggesting that hypercortisolemia is associated with hippocampal damage.

And a very recent study from the University of Michigan by Monica Startman where they

took 22 patients with Cushing's disease, looked at hippocampal volumes before and after transfenoidal adenectomy and they showed that there was up to a 10 percent increase in hippocampal volume in the same patients after the hypercortisolemia had resolved. So, again, suggesting that many of these structures are dynamic. So really depending on the intervals over which you measure the specific disorder in which you looked at these markers, they have to be interpreted accordingly.

Now methodologic issues, again, I'm not going to focus a lot on this particular slide, but it's important to keep in mind that there are many different techniques available to look at MR spectroscopy as well as volumetrics. And these have to be standardized across studies and, really, there's very few studies in the literature that have used the same technique.

For example, the acquisition protocols: What part of the brain are you looking at; what's the voxel size; how big is the volume element that you're looking at, and; really how are the data reported?

Are you reporting them by an internal normalization, by an external normalization, for example, to a phantom, are you atrophy correcting these data or are you reporting absolute concentrations? So these all some things that one needs to bear in mind and really standardize when you look at these studies.

So I want to summarize for you briefly the MRS literature in the Alzheimer's Disease. Now this slide lists the cross-sectional studies that have been done in the Alzheimer's and really the bulk of the literature is cross-sectional data. There are at least four postmortem studies that I could find with a total sample size of about 70 Alzheimer's patients, 69 Alzheimer's patients and 22 controls, mostly of the temporal and frontal cortex, and mostly based on per chloric extracts of postmortem brain. And they found a 20 to a 50 percent decrease in NAA in the regions of interest and a couple of studies have correlated this with plaque density. One study with plaque density and one study with tangles looking at it in adjacent sections.

Now, the in-vivo MRS studies, there's

about 30 studies or so. The sample sizes range from very small case series to more than 50 studies, more than 50 patients. The decrease in NAA in the Alzheimer's has ranged from about 10 to 40 percent, 10 to 37 percent with a couple of negative studies. In about four or five studies the NAA levels have correlated with many mental state examines with the Pierson. In small sample size studies you have a very high Pierson correlation and then the larger the sample size gets, your correlations tend to be a little bit lower.

Now there are two studies that have looked at the potential sort of prognostic role, if you will, of MR spectroscopy, and there may be more. These are the two studies I'm presenting today.

One was a study that we published, a pilot study that we did about four or five years ago where we looked at 12 very mild Alzheimer's patients, we did a baseline spectroscopy scan and then we evaluated them clinically over the next one year. And what we showed was that there's a correlation between their baseline spectroscopy measures and their cognitive

status one year later. We also found a correlation between the rate of change in their cognitive status as well, and that was also presented in this particular study.

Again, the correlation coefficients are high because the sample sizes are relatively small.

Now, the second study is from the Mayo group. Again Dr. Jack was one of the investigators in that. A study of 51 patients and the one analysis I'm showing you here is a pooled analysis they did where they combined the MCI and the Alzheimer's patients, so really this is the cognitively impaired group. And they looked at the predicted value of N-acetyl aspartate over Myo-inositol. Again, the ratios used sometimes because NAA goes down presumably in the Alzheimer's and MI goes up. So really one would expect this ratio overall to decline.

So this is a step-wise regression with age, education and various MRS ratios in the model. And this is the correlation that was explained, the predictive value of that MRS measure looking at various cognitive tests. This is the auditory verbal

learning test and this is the dementia rating scale. So really there is some predictive value for MRS measures.

If you look at longitudinal MRS studies, I could find only three studies with a total of 34 Alzheimer's patients and 14 controls. The follow-up in two of the studies was one year long, and in one of the studies was 23 months long. So that's the range of follow-up.

The methods varied. To my knowledge these were not controlled in these studies. I could be wrong, but the paper didn't mention it. In all three studies, in general NAA declined over time. The rate of decline was about 12 percent per year in the Alzheimer's and one percent per year in controls in the studies that reported a percent change.

The hippocampus decline in one of the studies that concomitantly measured hippocampal volumes -- I'm sorry, this was hippocampal NAA, this is gray matter NAA. The NAA and hippocampus declined 12 percent per year in AD, but it was not statistically different from that of controls in one

of the studies.

And in two of the studies the decline in NAA appeared to correlate with the cognitive decline.

So I want to present to you in the last few minutes a pilot trial that was done at Duke University, really to look at the effects of a cholinesterase inhibitor, in this case Donepezil on neuronal markers in Alzheimer's Disease.

Dr. Krishna is the principal investigator. The study, it's not yet published. And it was support by Eisai and Pfizer.

So this was sort of a Phase II study. It was a randomized double-blind placebo-controlled study of mild to moderate probable Alzheimer's Disease patients. MMSE score ranged 10 to 26. Twenty-four weeks of therapy with Donepezil or placebo. The Donepezil dose was 5 milligrams for the first month followed by 10 milligrams subsequently. And then after 24 weeks there was a 6 week placebo washout.

We obtained spectroscopy measures, MRI measures and the ADAS Cog every 6 weeks during the study.

We measured hippocampal volumes, but sort of this was a post-hoc analysis. This was not one of our A priori proposed outcomes in the protocol, but it was done in a blinded fashion, and it was only done at baseline and week 24. I think I'm not going to present those data just because if someone has a question on that, I'd be happy to talk about it.

The subjects were recruited at three sites, but all the scans were done at Duke.

So the trial outcomes, the primary outcome was N-acetyl aspartate, the secondary outcome was the ADAS Cog and other MRS measures. A post-hoc outcome was hippocampal volumes.

This is included in your slide set in your handout.

I'm going to show you some of the baseline characteristics. Really the baseline characteristics did not differ between the patients. There were 34 patients in the Donepezil group, 33 in the placebo group. You can see here the mean MMSE score is about 19. And these are the results on the ADAS Cog. The red line is the Donepezil treated patients, the yellow

line is the placebo treated patients. So these are the 24 weeks of the trial. This is week 30. And you can see that Donepezil, as expected, was better than placebo in terms of its effects on the ADAS Cog.

Now, we looked at a number of different regions of the brain, and I'm going to present to you the different regions in terms of our N-acetyl aspartate. Again, you can see here subcortical gray matter. You can see the red line again is Donepezil. That's placebo, the yellow. And, again, there is some inherent variance in the system and that's sort of reflected perhaps in that.

This is the cortical area, and the red line again is Donepezil. Now, our technique that we used was particularly bad for looking at cortical NAA because the voxel we choose cut out the rim of the cortex. So really there was a lot more noise in the cortex with this particular technique that we used at that time.

Now, these are the results for the peri/ventricular region. Again, you can see -- again, this is Donepezil. At endpoint really there was no

difference in week 24.

Now, this is the white matter. I just have a couple of seconds left, so I'm really going to finish with that slide.

And really I think our conclusions were that Donepezil improved cognition and increased NAA brain levels generally between weeks 6 and 18. However, drug-placebo differences were not significant at weeks 24 or 30. The variance was large. And really I think this was a pilot study that we did to try to come up with estimates of variant sample size, et cetera, and at least sort of demonstrates the feasibility, the technical feasibility of doing a study such as this.

So I want to thank you for your attention.

CHAIRPERSON KAWAS: Thank you, Dr. Doraiswamy.

Our next speaker is Dr. William Jagust.

DR. JAGUST: Well, thank you.

I would like to give you an overview essentially of PET and look at some of the reasons why PET is certainly interesting in this discussion, and

raise some questions perhaps for consideration.

So why should we consider PET potentially a good surrogate marker in Alzheimer's Disease? And I'm going to sort of outline my approach and then give you some examples.

So, PET is a reasonable good assay for tissue biochemistry and also for physiology that is intimately related to the fundamental disease processes of interest in Alzheimer's Disease. It's highly related to cognitive function. It's predictive of cognitive decline, very similarly to what we've heard about for MR and spectroscopy. And it is sensitive, reliable and reasonably valid as a marker of the actual pathology of AD, the amyloid plaques and the neurofibrillary tangles.

And, finally, it is statistically powerful and provides potentially powerful measures of disease decline.

Now, PET is actually a complicated technology. I think everyone understands that when we talk about PET, we're talking about a method of mapping in vivo radiotracers and what you label, and

the type of radiotracer we used, depends entirely on what you're interested in. I think for Alzheimer's Disease there are three potentially interesting types of radiotracers.

One are ligands that bind to cholinesterase and that reflect cholinergic function in the brain.

Another radioligands that bind to amyloid and in the last year we've hear more and more about this. These are very, very interesting types of ligands, but as yet I think we have to say they reflect to some extent unknown characteristics of amyloid and of the amyloid pathology, and they're not completely worked out in a number of ways.

And then what you'll hear about most today is fluorodeoxyglucose or FDG, the glucose metabolic tracer which all evidence points to largely, though not entirely, reflects synaptic activity.

As far as cholinergic ligands, some of the most elegant work on this was done by the group at Michigan who used this compound called PMP and showed that one can actually detect binding a cholinesterase

in the brain. This is a potentially very interesting technique that one could use to specifically assay the system and also measure effects of drugs that modulate the cholinergic system. I'm really not going to talk anymore about that, other than to point out that it's something that one needs to consider depending on the type of clinical trial you're interested in.

Now, this is an FDG, and the only point I want to make here is to show you the characteristic signature of Alzheimer's Disease on glucose metabolic studies, a controlled subject and two separate Alzheimer's patients both showing you an area of hypometabolism, posterioral here in the parietal lobes and also in the temporal lobe. There have been many, many studies that have replicated this, and many variations on it showing that it may asymmetric, it may be distributed slightly differently in different types of Alzheimer's patients, but in general this is the so-called metabolic signature of the disease that also extends into the posterior cingulate cortex, which in fact may be the most sensitive region of the brain for detecting early changes in Alzheimer's

Disease.

So, what about this in diagnosis? Well, there's been, again, in recent years more and more data gathering on how these metabolic patterns for glucose metabolism -- now again I'm talking about FDG PET -- relate to neuropathology, and I just picked two studies here. The first by John Hoffman and his colleagues showing that compared to pathological confirmed Alzheimer's Disease, this pattern has a sensitivity of about 90 percent and a specificity of about 65 percent for the diagnosis of Alzheimer's.

You'll probably hear more from Dr. Small, who?s talking after me, about this study, but this was a substantially larger study showing that PET was able to both predict progressive dementia in individuals who presented with cognitive impairment and also pathologically confirmed Alzheimer's Disease, in this case again with a fairly high sensitivity and specificity.

And so I think there is reasonable evidence from these types of studies that these metabolic findings are reasonably good markers for the

pathology.

PET also with FDG predicts cognitive decline, and there is really a plethora of studies that get at that particular issue. One study that we published a number of years ago shows that a baseline PET scan predicts a subsequent change in the Mini-mental state in patients with Alzheimer's Disease. Satoshi Minoshima's group, again, in Michigan showed that baseline PET predicts decline for memory impairment, or so called MCI to dementia, again showing changes in the cingulate were the most predictive of that type of decline.

More recently the group at UCLA has shown that baseline PET will predict memory decline in non-demented carriers who have the ApoE 4 gena type.

And finally there's been a recent study that suggests that PET may predict decline in normal individuals who go on to get a mild cognitive impairment.

So, again, I think ample evidence that PET can predict clinical course. And this is just an example of the study we published showing that at

baseline glucose metabolic rates predicted the subsequent change in mini-mental, those with lower metabolism declined more rapidly over the ensuing two years. And this actually remained significant when one controlled for a number of demographic factors.

Here showing you that glucose metabolism is related to cognitive function in the sense that what we see here on the Y axis is a memory performance and on the X axis a glucose metabolic ratios in the temporal lobe and in the hippocampus. Just an illustration of another finding that's been fairly widely documented that particular types of cognitive deficits are correlated with regionally specific patterns of glucose metabolism.

Now, I want to talk a little bit about progression and change, and measurement of change over time. And I'm going to rely on data that was published by Eric Reiman when he studied a group of individuals who were asymptomatic who were ApoE 4 heterozygotes with repeated sequential PET scanning over time. And what you see here is the change in glucose metabolism or the decline in glucose metabolism over a two year

period in these individuals.

And one can look at this quantitatively and simply look at the normalized, in this case the region normalized to hold brain glucose metabolism over time. And one again sees decline over time. And using these kinds of data one can begin to look at the numbers of subjects one needs in a clinical trial. And Dr. Reiman published these figures in his paper.

And one can that depending on the size of the drug treatment effect, and this here represents the size of the change in glucose metabolism one was postulating, you would need relatively small numbers of subjects who are ApoE 4 carriers to detect a change of this magnitude using posterior cingulate glucose metabolism with 80 percent power.

If one looked at ApoE 4 noncarriers, these numbers got slightly larger, but still are in the manageable range. And in fact, when one looks at actual patients with Alzheimer's Disease to detect a treatment effect, one sees that even with a very small treatment effect on patients with Alzheimer's Disease, the number of subjects one needs for a clinical trial

of this sort is actually quite small. This year projected using frontal glucose metabolism, again with 80 percent power.

So statistically, at least, this is a manageable approach if one is convinced that measuring this size of reduction in glucose metabolism is what's necessary.

So, let me sort of philosophize about this now. Because this is where the data meets the road, and maybe we don't know how that's going to work out.

So, here are the positives about, I think, FDG PET as a surrogate marker. And I think that largely relates to the side on linking PET scanning to clinical declines or to the clinical side of the disease. And that is, as I showed you, PET predicts clinical decline and prediction, we understand, does not make a surrogate.

Also PET is biologically plausible. We've heard that word a lot today. It may well be on the disease pathway, and I think there are several reasons for believing that it might be.

The first is that it is reasonably

sensitive and specific for the pathology, for the way we define Alzheimer's Disease, for plaques and tangles, that temporal and parietal glucose metabolism seem to reflect that.

It's related to synaptic function. Glucose metabolism, largely related to that. And we increasingly believe, I think, that synaptic dysfunction is a key component of the pathological process in Alzheimer's Disease, and it's correlated with cognition. And, of course, it's statistically powerful.

But the negative, and the question that's been raised, I think, subtly and really needs to be discussed clearly is what is the link between using PET and trying to detect an effect on a disease that's underlying modifies its progression. And that relates to the question of whether PET can distinguish symptomatic therapy or state effects from underlying disease modifying, drug effects. And there's no easy answer to this.

Obviously, the one that's been proposed for clinical trials is to use a randomized start or

withdrawal design. Another, I think very important thing that we need to be considering is the use of PET tracers that really reflect the basic biology of Alzheimer's Disease. And to the extent that that may be amyloid, the PET amyloid imaging agents really offer a tremendous option in that direction.

The other point that I want to make is that state effects, as far as we understand them at least for cognitive states, are relatively small compared to disease effects. I showed you a PET scan of an Alzheimer's patient, you could see that that Alzheimer's patient's scan looked different than the normal control individual.

Individuals performing cognitive tests have metabolic rate changes on the orders of several percent. You can't see that in an individual image, which is why subjects are averaged over numbers of studies.

So cognitive state effects are relatively small. Disease effects, cognitive effects, several percent disease effects, 20 to 30 percent.

Drug effects are unknown, and that's I

think still an unanswered question.

But sitting and listening to this I was maybe naively wondering if this isn't a subset of the larger problem, which is that when we start to talk about surrogate endpoints and clinical outcomes being on different pathways. And really, let's take the perhaps trivial but nevertheless potentially important issue of fluid balance in MR. I mean, that's an effect that's going to change a surrogate endpoint, perhaps it has nothing to do whatsoever with the clinical outcome we're interested in.

Suppose we have a drug that has a direct effect on glucose metabolism. All it does is increases glucose metabolism. An amphetamine, for example. That's the same kind of problem. And I think what this says to me is that when we're thinking about symptomatic or state effects, we really have to understand the effect of the drug on the surrogate that we're measuring, just the same as we have to understand how a drug affects fluid balance if we're going to make measurements of MR atrophy. We have to understand how a drug affects glucose metabolism

independent of its effects on Alzheimer's Disease if we're going to use glucose metabolism as a surrogate marker.

So just to sort of make a couple of last points about technical issues. These studies, as are MR studies, can be technically very complicated. And there many issues that need to be considered in designing a multi-site acquisition study. That is, of course, subjects? state, other drugs they may be taking and so forth. How one is going to quantify the image which particularly involves whether one is going to measure the input of the tracer to the brain, which in a truly quantitative study requires a catheter in the radial artery, but there are alternatives to that. And how one is going to measure attenuation and then differences in instrument resolution across sites.

And then standardization of data analysis is the flip side of all this where one can quantitate these data with metabolic rates or ratios and whether one chooses region of interest or atlas or voxel-based approaches. These are very complicated, but I believe they're all actually manageable.

So to summarize, there is no doubt that PET is not a confirmed surrogate. That's a very easy question to answer. I think some of the data, some of the data that I showed you and some of the evidence, suggests that it has a lot of potential in that direction. It's sensitive to decline, statistically powerful, it has strong links to clinical symptoms, to pathology. But there are real questions about its relation as a disease modification marker. Any clinical trial, as I said, has to assess potential state effects on the PET tracer of interest, no matter what the PET tracer is. And I would submit, no matter what the imaging modality is.

And really, I think the only way we're going to answer many of these questions is if we begin to cooperate PET in clinical trials when we finally have disease modifying drugs. That's the only way we're really going to get at answers to a lot of these questions.

So, thank you.

CHAIRPERSON KAWAS: Thank you very much.

And our final speaker for this portion of

the meeting is Dr. Gary Small.

DR. SMALL: Just getting the technology to communicate here. I'll take a second.

Well, thank you. I'm delighted to be here and have a chance to expand on some of the comments that Dr. Jagust just made and throw in a few of my own in my discussion of positron emission tomography in dementia.

I want to start off with the point that PET is an imaging technique that provides information not just on brain structure, but also on the biochemical bases of brain function, which to me is importance since we're looking at in terms of response to drug treatment. And we would expect that most drugs would have an effect on biochemistry of the brain.

As we just heard, many of the studies have involved glucose metabolism using 18-F-fluorodeoxyglucose, which demonstrates the specific patterns of cerebral metabolic metabolism in various dementias. And there's extensive work in this area. In fact, we have about 25 years of experience. And I've just listed some of the many studies that have shown

some of the patterns that we've just heard about.

As we've seen in early Alzheimer's Disease, parietal regions, temporal and even frontal regions begin to show this deficit that progresses to late stage Alzheimer's Disease. And interestingly, late stage Alzheimer's Disease has a pattern that looks very much like an immature brain, as we see in this image.

We also see different patterns in different types of dementia. Here, again, is an Alzheimer's case with the parietal hypometabolism, a vascular case with both cortical and subcortical deficits, frontal dementia or Pick's Disease with frontal hypometabolism, and the caudate hypometabolism in Huntington's disease.

Now, last year Dan Silverman led an effort, an international effort, to look at the regional brain metabolism and long-term outcome with PET. And I've listed all the many collaborators, many of them are here in this room, and involved centers throughout the United States and Europe. And we asked questions such as we see in this slide, what is the

accuracy of FDG-PET for assessing the presence or absence of a neurodegenerative dementia. So we, as in this column, neurodegenerative disease as seen on PET by blinded reading and then in this column or this row neurodegenerative dementia present on autopsy. And with these kinds of numbers one can calculate the sensitivity 94 percent, specificity 78 percent, and overall accuracy 92 percent.

We can ask the question presence or absence of Alzheimer's Disease, we see similar kinds of sensitivities and specificities. And that was on a sample of about 130 patients who were followed up to autopsy.

On another group we followed at least two years on an average of about three years, we asked how does PET predict the progression of dementia. And we saw similar results in terms of sensitivities and specificities.

So our conclusion from that study was that Alzheimer's Disease and other progressive dementias significantly alter brain metabolism early relative to the manifestations of clinical symptoms. And the

clinical FDG PET detects this altered metabolism providing an accurate clinical tool for noninvasive prognostic and diagnostic assessment.

And if one looks at studies where we use conventional clinical assessments, where we have repeated examinations not using PET, we get lower sensitivities and specificity. So in the study that has 134 patients with autopsy criteria as the outcome where multiple examinations were done over the course of several years, we find lower sensitivities, around 83 to 85 percent and lower specificity is about 50 to 55 percent.

So these data suggest that PET is a reasonably valid marker of clinical progression and of autopsy findings.

Now, we saw in other material handed out that one of the problems with the specificity of diagnosis not using PET is in differentiating frontal temporal dementia from Alzheimer's Disease. And at the International Alzheimer's Congress in Stockholm, Norm Foster presented some data that I thought were quite interesting, specifically looking at this question,

blinded assessments, very well controlled study again, involved several sites and they had very high inter-rater reliability among the raters and high diagnostic accuracy, about 80 to 90 percent in just differentiating frontal, temporal and Alzheimer's type dementias.

For the last several years we've been looking at how well PET performs in detecting very subtle brain changes in people without dementia, people maybe in their 50's or 60's who have just minor memory complaints. So we've been studying middle aged people with the genetic risk for Alzheimer's Disease, apolipoprotein E, or ApoE 4. And back in 1995 in 1995 we first reported that you could see these changes. Eric Reiman's group at the University of Arizona has rated those findings. And both our groups have in independent samples published additional data and also data showing how there is change over time. And I'll just show you some of that information.

This is a study we published a couple of years ago in PNAS where we had 54 subjects, half of them had the genetic risk for Alzheimer's Disease, the

other half did not. They all had very minor memory complaints. On the average, they were in their mid-60's. And the statistical parametric map shows you where in the brain there was significantly lower metabolism in people with the ApoE 4 genetic risk. So the lateral temporal, parietal, dorsal lateral prefrontal cortex and posterior cingulate cortex had these changes.

When we followed people with ApoE 4 over a two-year period we found that these same regions, the parietal and temporal regions, showed decline. About a 4 to 5 percent decline in these critical brain regions. This is just in ten subjects, and you can see there's no overlap from baseline to follow-up in this right lateral temporal region in terms of the metabolic decline.

Now, as we just saw, based on those kinds of data we can begin to make power estimates of how many subjects we'd need in a clinical trial to be able to show a treatment effect. And we saw some of these data just a moment ago. Just to summarize what the model looks like, instead of looking at cognitive

function, if we looked at metabolic function in critical brain regions in an ApoE 4 subject on placebo, one would expect this decline. About 4 or 5 percent over a two-year period. And if the active drug is working, we would expect a slower decline.

Eric Reiman's group has extended these studies to look at patients who already have clinical dementia or Alzheimer's Disease. And these are figures taken from his article this year showing the significant differences between patients with Alzheimer's Disease and controls. The areas where there is lower, significantly lower glucose metabolism. Again, parietal temporal regions, posterior cingulate regions. And he has followed these patients over time. And here we see the areas where there is significant decline in these same brain regions.

So to summarize, if one were going to study FDG as a surrogate marker in brain aging clinical trials, if we had a drug with a 33 percent treatment effect in the pre-symptomatic cases if we're just going to study ApoE 4 subjects, we only

need about 60 subjects per treatment group. And this would be based on a two-year study. If we're looking at patients with Alzheimer's Disease, we'd need an even smaller number over a one year period, 36 subjects. And in those studies it's best not to stratify according to genetic risk.

Now, what is the experience thus far with treatment trials looking at PET changes? And I was able to find in the literature and also just in press three studies that I think are relevant for this discussion.

And also while I'm on it, let me mention other conflicts of interests that weren't mentioned. That is that I have been an advisor in the past for Bayer and have advised Novartis, Eisai and Pfizer as well as Janssen.

So in this study from our group we looked at the cholinesterase inhibitor drug Metrifonate. It was a 6 to 12 week treatment period, a relatively small number of subjects. And we found that there was cognitive improvement in all of the subject, at least a 2 point increase on the Mini-Mental State Exam, and

also significant increases in glucose metabolism, particularly in these key regions I've been talking about, parietal, temporal and frontal.

Steve Potkin at Irvine headed up a study of Rivastigmine, and this was a 26 week double-blind study, placebo controlled, 27 patients in this study, and they showed very interesting results. 33 percent increased in hippocampal metabolism in the responders, those who responded to the drug who had clinical improvement. But the non-responders show a 6 percent decrease in hippocampal metabolism, which was similar to what was seen, the 4 percent decrease in the placebo treated patients.

Another study that Laurie Tune has presented at several meetings and now is in press, is a study of Donepezil. Again, a blinded study, 24 weeks of treatment. And they found that mean glucose brain metabolism remained stable in the active drug group and declined 10 percent in the placebo group. And there were significant parietal, temporal and frontal treatment differences in the study.

Now, here is an image from the study we

did with Metrifonate showing average PET scans before and after with Metrifonate. And here you can see, particularly in the parietal regions, this is pre-treatment and post-treatment where there is that increase in metabolism. And at a lower level in the brain you can see an increase in frontal regions as well as some of the temporal regions.

In the study with Donepezil, this is showing the Donepezil and placebo treatment effects on relative average glucose metabolism. This represents placebo and this is Donepezil at 12 weeks and after 24 weeks you can see the placebo group declines but there's stabilization in the active drug group.

Now, when we think about PET multi-site trials, I just wanted to cite a point made from our Alzheimer's Association Neuroimaging Work Group, the PET Research Subcommittee. And we talked about some of these methodological requirements that we've heard discussion of. Also that there should be consideration given to trials where we include both PET, FDG and MRI because of the different kind of information involved. And also an interest in PET

radiotracer methods that will image the pathologic lesions. So I wanted to spend a few moments talking about this.

This has been in the news lately and there have been various approaches. One approach is to alter conventional dyes used at autopsy such as Chrysamine-G and they're effective in vitro, but they don't seem to cross the blood brain barrier.

Now University of Pittsburgh has pushed this approach forward, and actually did develop a probe that crosses the blood brain barrier. We saw some of the limited in vivo data in Stockholm. They've scanned about a dozen subjects. One of the limitations thus far with that, it uses carbon 11 as a labeling probe and that has a 20 minute half-life, and is a bit awkward in clinical settings.

At UCLA we've been developing what we call FDDNP and we've shown that it's effective both in vivo and in vitro. We have fluorine 18 labeling, so it's much easier to use clinically. It has 110 minute half-life.

We have information on over 60 human

studies that we've completed to date, and we're in the planning stages of multi-site studies. And we also have postmortem neuropathological validation of our in vivo data.

DDNP is a fluorescent small molecule probe. It's neutral and lipophilic, and it was originally developed for fluorescent microscopy. And as we'll see, it provides excellent visualizations of neurofibrillary tangles, neuritic plaques and diffused amyloid.

We call it DDNP, it stands for dimethylamino dicyano naphthalenyl propene and our chemist George Barrio adds fluorine 18 at this end of the molecule. If one looks at time activity curves and you plot radioactivity versus time, you can see there's very good uptake in the first 10 minutes. And after about 30 or 40 minutes, one sees the signal here where in temporal regions there's a greater retention or activity of the molecule compared with other regions. And here you see in the temporal region the increased activity.

So, if we look at a patient with

Alzheimer's Disease, this is an MRI scan, you can see the atrophy by the increased ventricles, this is an FDG PET scan showing lower activity reflecting lower neuronal activity in temporal regions and the DDNP scan shows higher activity reflecting what we think is a greater accumulation of plaques and tangles.

We've plotted the signal against various cognitive measures. And with the Mini-Mental State Exam and you see a good correlation in controls as well as patients. And we've done similar studies with more sensitive memory scores, such as the immediate paragraph recall score and the delayed paragraph recall score you see very high correlations with the signal. It separates patients with Alzheimer's Disease very well from controls. And this is the postmortem study I was talking about here. You can see a coronal section of temporal activity. The patient died 8 months later, and this is autoradiography showing temporal and parietal activity superimposed on the in vivo scan. And the inset shows you confocal microscopy of plaques and tangles.

We are just now studying other kinds of

dementia. This is a scan of a patient with a clinical diagnosis of frontal temporal dementia showing you activity in temporal regions as well as frontal regions. This is the FDG PET scan showing you a slightly different profile.

And I think the great strategy that we've been alluding to is to include multiple sources of information in these kinds of studies and also to ask what kind of question is important. If you want to look at neuronal function, FDG PET is a good marker. Plaque and tangle load, DDNP PET. We want to add information about genetic profiles and neuropsychological functioning, and there are a variety of other approaches; structural imaging, MRS, functional MRI that can add additional information.

So, in conclusion I just wanted to mention or review some of the points made by our neuroimaging work group of the Alzheimer's Association, including myself, Norm Foster, Bill Jagust, Eric Reiman and Moni deLeon. We thought that PET compliments structural imaging, it can serve an in vivo biomarker to improve clinical care and research in Alzheimer's Disease.

It's clearly becoming increasing available. It can confirm the presence of a neurological disease in mild dementia and assist in differential diagnosis.

We felt that it should be considered an option for the clinical diagnosis of Alzheimer's Disease. It shows potential for predicting prognosis in people at risk for dementia and assisting in new treatment evaluation, increasing efficiency of prevention therapy, testing, increasing understanding of dementia diseases.

Randomized multi-site clinical trials are needed to further assess clinical applications and its use as a surrogate marker in drug development.

Alternate methods of data analysis need to be compared, and the most effective one standardized.

Development of new PET ligands, we strongly recommended. And we felt that PET should be included in all clinical trials where Alzheimer's Disease is sought as a pathological substrate for the therapy.

And just to acknowledge some of my many collaborators and many funding sources.

Thank you very much for your attention.

CHAIRPERSON KAWAS: Thank you Dr. Small. And thank all the speakers in this section for their informative presentations.

The floor is now open for questions to the presenters on MR spectroscopy and PET.

I'll start then. I'd like to ask Dr. Doraiswamy, the MR spectroscopy data that showed NAA wash-out by week 24 and 30, what do you think is going on there?

DR. DORAISWAMY: Don't know the answer to that. It would be nice if we had another drug. I mean, right now drugs like Donepezil are what we consider the gold standard for treating Alzheimer's. If we had a true disease modifying drug, for example, an antiamyloid secretinase inhibitor or something, and if we compared the two, then we would have a true sense for what would happen with an antiamyloid drug.

The second thing is in this particular study at weeks 24 and 30, you're getting some subject attribution. It's a very small sample size study, so it's really hard to tell if what we're seeing there is

a true lack of effect or is it really a sample size effect. So I really can't answer that.

At the time we planned the study we didn't have any good data to estimate sample sizes for this kind of a trial and we based it just on the logistics of, you know, doing a small pilot study.

CHAIRPERSON KAWAS: Thanks.

Actually, I have a second question for anybody who presented.

We saw intriguing data on drugs? effect on various modalities. Has anyone ever tried using a non-AD drug to make sure that we won't get the same effect, a drug that we don't believe should be affecting Alzheimer's Disease that we're sure it doesn't make similar changes? How specific is the effect, I guess is what I'm really asking?

DR. JAGUST: Well, I think, for PET any drug that has an effect on glucose metabolism will affect the results. So, you know, I alluded in my talk to amphetamines. You know, I mean an amphetamine or barbiturates, I mean, they don't have a fundamental effect on Alzheimer's Disease. They may change

patient behavior, but they'll certainly change glucose metabolism.

So I think, you know, my point is that anytime -- any drug can have an effect on the kind of signal we're looking at in a PET scan, and you have to understand what the underlying physiology is in order to interpret the images.

CHAIRPERSON KAWAS: Thanks.

DR. DORAISWAMY: I have a comment. I think in mild to moderate Alzheimer's Disease it may not matter as much as in the advanced stages where people are taking anti-psychotic drugs and there's evidence form the anti-psychotic literature that some of these drugs could have potential effects. So at that point, again, that's a very good point. People may need to control for anti-psychotic use.

CHAIRPERSON KAWAS: Yes. Okay.

Okay. Dr. Wolf has a question.

DR. WOLF: Yes. My question probably is directed more generally because coming from the imaging side and not from the neurological side, I'm not that up on what is a mechanism of Alzheimer's.

I would like to know to what extent is the amyloid plaque deposition then reflected in intracellular changes? Because what we see in the case of N-acetyl aspartate and FDG are all events that happen at the intracellular level whereas my understanding is, and I stand to be corrected, is that the amyloid plaque are at the extracellular level and therefore affect somehow what gets into the neurons or not.

So the question is do we have any measures on changes from the spectroscopy either from the spectroscopy or from the PET that tell us the rate of change or the measures of amyloid plaque that could direct us then to what is happening in that disease?

CHAIRPERSON KAWAS: Does any of our invited speakers want to tackle that one?

DR. SMALL: Well, I don't want to tackle it, but I've got the microphone, so I'll try to address it.

The amyloid plaque correlates with the disease. And, you know, whenever I talk about the DDNP, plaque and tangle imaging, I try not to get into

debates about the amyloid hypothesis. It may be that the DDNP would be a great way to track plaque and tangle or plaque deposition. And the good news at the end of the day to the patient might be you have no plaques in your brain, your DDNP scan looks great. And the bad news is the patient wouldn't remember the conversation.

So, you know, whether it's correlated with the disease or not, I don't -- whether it's actually -- if you can clean out plaques from a certain drug, you still may not be able to cure the disease. But I think the point here, and actually with all these markers, is getting back to the critical question, is it a good surrogate marker? Does it correlate with clinical progression? And if it does, is it something we ought to be measuring just like anything else and leave it up to the drug trials to prove or disprove a particular underlying path of physiological mechanism for the disease.

DR. DORAISWAMY: The only thing we know from spectroscopy is that NAA appears to decline over time in areas that are effected progressively by

amyloid. To my knowledge there's no studies in any of the animal models of Alzheimer's Disease, even though there are studies in animal models of ALS and other conditions. And there's only two postmortem studies that have correlated with amyloid, and they're small sample size studies. So that's the amount of the information we have.

DR. JACK: You know, it's my understanding that the toxic agent in fact is oligomeric fragments, so beta amyloid oligomeric fragments. And in that sense it may be that there is no perfect biomarker for Alzheimer's Disease. The biomarker for the disease in fact is a measure of the abnormal metabolism that over-produces these oligomeric fragments.

And so every marker, even direct image amyloid burden, in fact may turn out to be a somewhat of an indirect market. So the same limitations that apply for markers that everyone admits are indirect, glucose metabolism, brain atrophy, NAA, et cetera may in fact apply to direct measures of amyloid load.

DR. SMALL: I just want to clarify one thing and then make another point. And that is the

Alzheimer's Association Neuroimaging Work Group information that I mentioned, this was information that was reported at the Stockholm meeting. And it is a work in progress. The entire committee is still going over this information. So if anything, it reflects the opinions of just the subcommittee, the PET subcommittee, and it's still being edited and worked on, so it's not an official position. I just want to clarify that.

The other thing is in this discussion, you know, it seems to me that many of the arguments made about MRI also fit with PET. I mean, we're talking about disease modification. The way to determine disease modification would be these delayed-start study designs and similar study designs. Because one may find you make -- for example, you give a drug to somebody and let's say hippocampal volume increases; you take the drug away, that volume increase may go down. We just heard about the example with alcohol.

So just the fact that a structural change occurs doesn't prove that it's a disease modifying -- and the same thing is true about FDG PET. I mean, I

showed you some data where we saw increases in metabolism with cholinesterase inhibitor drugs. I didn't give any information what happens when we withdraw the drug. We're presuming that it's going to be symptomatic just as we see with the clinical data. But it is possible that a drug could produce a disease modifying effect and you could see that on a functional image. You could withdraw the drug and you could still see improvement in neuronal function.

CHAIRPERSON KAWAS: Our last speaker for this session is Dr. Michael Hughes, who?s returning to talk to us about validating surrogate endpoint.

DR. HUGHES: Thank you.

I'm going to pick up where I left off this morning. I'm really going to focus not on biological models, but looking at empirical evidence from studies to support the validation of a surrogate. And what I'm going to do is illustrate the talk a little bit with experience from HIV where there was a collaborative effort to validate viral loads? surrogate endpoint. And that's now actually been incorporated into a recently released FDA guidance on that issue.

From a statistical perspective, the most commonly cited definition of a surrogate is really framed in the context of hypothesis testing. And more importantly, this criterion gives rise to two operational criteria which are sufficient for validating a surrogate endpoint.

The first one really deals with the issue of correlations, so whether it's a prognostic marker or not. And the second the deals with the idea that the surrogate must fully capture the net effect. By net effect, it means combination of adverse and beneficial effects of treatment on the clinical outcome. And as I mentioned earlier, both are required. Correlation itself is not sufficient.

This second criteria, it really fits very well with the part of the Temple definition about establishing that changes induced by therapy on a surrogate are expected to reflect changes in the clinical endpoint.

So you can develop a framework which might used for establishing surrogacy. Firstly, the surrogate must be a prognostic marker so you can deal

with that in natural history studies.

Second, is that treatment mediated changes in the surrogate must be prognostic. And that requires interventional studies.

And the third is whether the effects of treatment on the marker explain or are associated with the effects of treatments on the true clinical outcome.

So I'm going to talk a little bit about the second one, then come back to the third one.

Here's an example which I hope will show you that just looking at early changes is not -- treatment mediated changes are not sufficient for validating a surrogate. So here's a typical situation where subjects are classified as whether they respond to treatment or not, yes or no. And you can see this is an HIV example that the responders near the bottom here had a much lower rate of progression to AIDS or death than the nonresponders. And if you look at this quantitatively, it's highly significant.

But in this particular example I've chosen, this was a placebo treatment. And it really

opens up the possibility that healthier subjects could respond to the therapy that you're studying.

So you cannot establish that a response variable was a good surrogate using data from an observational study of treatment mediated changes or a single arm of a clinical trial. However, it's important that the association between the treatment mediated change in the surrogate and the clinical outcome doesn't depend upon the intervention. Clearly, if it depended upon the intervention, then when you go to a future study you don't know how to interpret the results, the marker results, surrogate endpoint results.

Here's an example of what was done in the collaborative projects. So this is a plot which shows for a large number of clinical trials and for a very broad range of treatments within a particular class of treatments, shows the estimated association for a one log reduction in viral load and its association with progression to AIDS or death. And the fact that most of the estimates to the left of the line shows that reducing viral load using these treatments is

associated with better clinical outcome.

And if you look at the very bottom right hand corner, there's a test of heterogeneity which establishes that from this data there's no significant evidence that association varies between the different interventions studied.

And here's a similar one for a CD4 cell count.

So let's go on and think about the third aspect, and that's trying to establish that there really is an association between the changes that are induced by a surrogate endpoint and the changes in a clinically meaningful endpoint. And it's useful here to remember that the real way to show that a treatment induces changes in outcome is to use a randomized trial of that treatment.

And so one must ask the question then how can we use information from the randomized trial to validate a surrogate? So I'm considering a hypothetical trial which is comparing treatments A and B.

A single trial in itself is most useful

for providing evidence against surrogacy, as the case studies this morning showed. Clearly, if you get the effects on the clinical outcome and the effects on the surrogate going in opposite directions, then that's evidence against surrogacy.

If you have a very well powered trial for the clinical outcome which shows very similar outcomes, but you find a significant difference in the effects on the surrogate, and again that's useful evidence against it being a good surrogate.

Having said that, the interpretation of this sort of information from a single trial really needs to be set in the context of a large number of clinical trials and assessing whether this happens very rarely or is a common problem.

So what can be done when you've got effects going in the same direction, so the effects on the surrogate and the clinical outcome are pointing in the same direction? Well, the first thing that might be asked is whether the association between the surrogate and the clinical outcome varies between the randomized arms. In other words, whether it varies

between the interventions being studied.

And if you find what statisticians call a significant indirection, in other words the association between the treatment mediated changes and the clinical outcome varies between the interventions, then that's evidence against surrogacy. It means it's not going to be reliable for future studies.

And clinically what this really means is that the way that you interpret the different changes for individual patients depends upon the specific intervention that was used to obtain those changes.

The next thing I would like to talk a little bit about is the idea of what people call the proportion of treatment effect explained. And this really came -- this idea came out of Prentice?s second criterion that a perfect surrogate must fully capture the net effect of treatment on the clinical outcome. And in an imperfect setting we're not interested in fully capturing it, but in partially capturing it. And there is this concept to proportion of treatment effect explained, which is in the literature and has been used, but is now largely discredited.

And the reason for this is that the notion of a proportion here is fallacious; that you can actually obtain values outside of the range of zero to 1. So finding a proportion of one doesn't mean you've necessarily got a good surrogate. It explains the treatment effect on the clinical outcome.

So in terms of what you can do in single randomized trials, I think the most beneficial use is actually providing evidence against surrogacy.

In terms of evidence in favor of surrogacy, generally I think the opinion is that the framework there is somewhat flawed. And I personally think it's very unlikely that any method will ever be useful in a single trial because what you're trying to do is explain a treatment difference which generally is imprecisely estimated in the first place. So your ability to explain it is always going to be weak.

So, the obvious step then is to go into a meta-analysis of randomized trials. And I think this is the approach which is more broadly accepted now.

The basic idea here is to evaluate the association between the difference in effect on the

true clinical outcome. So the difference between randomized arms and the corresponding difference in effect on the surrogate across multiple trials. And it's important to appreciate this uses information from all trials so you have the standard issues of meta-analysis about trying to obtain information from all available trials that address the question of interest.

And this is a schematic of what you're trying to get at. So you can imagine each of these points, the center of the cross, being an individual randomized comparison. So we're asking is there a correlation between the differences between randomized arms in terms of the clinical outcome and the differences between the randomized arms in terms of the marker outcome.

And so we've got a large number of randomized comparisons here. And the arrow bars are meant to just make the point that within any individual trial, as in precision and both estimated in the clinical outcome as well as the marker outcome.

So this would be a schematic for a good

surrogate endpoint. So if you imagine in a future trial you estimate a marker difference up here between two treatments, and then you can imagine drawing a line down and then across and you could get an estimate or a prediction for what might be the likely difference in clinical outcome.

And this is a similar schematic of exactly the same situation I've just shown where instead of using arrow bars, the size of the circle represents the amount of information coming from the trial. And it tends to show the association somewhat more clearly.

Now if you have imperfect surrogates, then the effect of that is usually to -- or will be to produce a more diffused association or even no association.

And if you have an intervention which has an adverse effect on the clinical outcome, what it will do is produce points either in the upper left quadrant or the lower right quadrant here reflecting the difference in direction of effect.

So this collaborative group did this, and

obtained data from all randomized trials of one particular class of treatments in HIV. And the markers of interest were a measure of viral load and a measure of immune function. And the true endpoint was what was typically used in clinical trials at the time, which was progression to AIDS or death.

I think a key thing here is that this was a very successful collaboration between pharma and academia in obtaining very extensive data. I don't think there was a single trial that was missed in this meta-analysis.

And this shows the situation for viral load. And I think the most important thing is in the two quadrants, the top left and the bottom right, there are essentially no points or points with very, very little information. So there's no real conflict between the viral load results from these trials and the clinical outcome results. And, in fact, statistically you can fit a regression line through this and you find evidence of surrogacy.

CD4, it's actually more impressive except that you've got this one trial which is clearly having

inconsistent results between the marker and the clinical outcome. And this, perhaps, isn't particularly unexpected in that CD4 is a more proximal outcome to the true clinical outcome than viral load is.

So, I thought I'd finish just by summarizing what I thought of some of the issues facing the validation of surrogate endpoints in Alzheimer's Disease. I think there's a key issue here about what is the true clinical outcome that needs to be considered. Clearly you're looking at an association between effects on a surrogate and effects on a clinical outcome. And if there's multiple clinical outcomes, then you want to look at multiple possible associations.

You really do need some sort of systematic evaluation of the prognostic value of treatment mediated changes, so that means going into your trials and looking at whether the changes in the markers really predict the changes in the clinical outcomes.

And you want to ask yourself does this prognostic value vary much between populations and

more particularly, between different interventions.

To be honest, I think the biggest challenge of doing anything like this is just getting people to share data and undertake this sort of systematic evaluation, whether it's done at a qualitative level or at a very quantitative level. But I think this is a key issue. And, obviously, the lack of large numbers of longer term trials at the moment in Alzheimer's Disease also limits your ability to do this. But this collaboration is really the essential facet of being able to validate a marker.

Thank you.

CHAIRPERSON KAWAS: Thank you, Dr. Hughes. That was very informative.

We do have time for a question or two before lunch break. Dr. Katz?

DR. KATZ: Yes. I think we use the term "surrogate" in a number of different contexts. One important use is whether or not a surrogate has been validated so that in the next study one could only look at the surrogate and not have to worry about looking at the clinical.

In the other sense, people have already started to, in a preliminary way, talk about the utility of using imaging in conjunction with a clinical outcome in a particular trial of a particular drug and suggesting that if those are both correlated in a single trial, or maybe if it was done twice, with the same drug, that that would support a claim that that drug specifically had an effect on progression. So not so much interested in using the surrogate in the former sense in which I just discussed, in other words not so much worrying about whether or not that that surrogate can then be used with other drugs, but just for that one drug if there's a correlation in a given trial or in two trials between a clinical outcome and the surrogate.

In your view, would that sort of an outcome support a claim for that drug for an effect on progression?

DR. HUGHES: People have looked at this issue in other diseases. And the basic idea and the way it's been used in other diseases is to model the association between the marker and the clinical

outcome within each of the randomized arms of the study. And then use that model to try and boost your precision in estimating the difference in the true clinical outcome.

And it's been used with, I have to say, very moderate success. The gains that you get from a statistical perspective, in other words the gains in precision, are usually quite minimal. And that's because the model describes the association between the marker and the clinical outcome is often not very precisely estimated.

So, I think you can from a statistical point of view you can use the joint information to support the licensure of a single drug. But I don't think the gains that you'll obtain within a single trial are going to be particularly marked.

It's really driven at the end of the day by the information that you've got about the difference in clinical outcome between the randomized arms. And usually that's very imprecise.

CHAIRPERSON KAWAS: Dr. Van Belle.

DR. VAN BELLE: Just a question. A lot of

the information we heard this morning deals with non-randomized observational studies. Any role for observational studies in evaluating the effectiveness of markers or surrogates?

DR. HUGHES: Well, I certainly think observational studies are very important. I think they have a definite role in establishing that the marker predicts the outcome, clinical outcome. I think you can use data from observational studies to establish that changes in marker levels that follow the initiation of a treatment also predict changes in outcome. But there's no way that you can use observational data to really establish that a marker is valid in the sense of drug approval. In other words, you can never fully establish that the marker effects explain the clinical effects. You've always got this possibility of an association in the healthier subjects may be the ones that respond to the therapy.

So I think you do ultimately have to go into randomized clinical trials to get the final piece of information that you need.

CHAIRPERSON KAWAS: A final question from Dr. Fogel.

DR. FOGEL: Yes. I have two really quick things.

One was you said on one of your slides that if the surrogate goes in opposite ways to the clinical outcome, that that is against surrogacy. But if they reliably go in opposite directions, couldn't that actually be used as a surrogate since you know if it's going one way, you know the clinical outcome is going to go the other way?

And I guess the second question I had was in parentheses, second sufficient condition that the endpoint must capture the net effect of the treatment on the clinical outcome, which meant beneficial as well as adverse effects. Are we in danger of actually throwing away good surrogates because the adverse effects may be drug specific and that in other drugs where the adverse effect may not be there, you could still use it as a surrogate but it's just because it had adverse effects that were specific to that drug that you've thrown it away?

DR. HUGHES: To answer your first question, in theory any marker that reliably predicts the clinical outcome could be used, even if the marker goes in the wrong direction. However, that's a statistical answer and I think it's really critical, though, that you have an underlying biological model which associates how the marker should behave with the clinical outcome.

In terms of your -- I'm sorry, I forget the --

DR. FOGEL: The second question, about second criteria for a surrogate.

DR. HUGHES: And what aspect?

DR. FOGEL: And whether or not you might be in danger of throwing away a good surrogate because the adverse effects may be specific to the drug and not because it's a bad surrogate.

DR. HUGHES: Sure. I think that's a good point. And what it really it emphasizes is the need not to look at a single study, but to look at multiple studies involving different drugs. And you're really interested in establishing consistency across a range

of interventions.

CHAIRPERSON KAWAS: Dr. Love?

DR. LOVE: Just semantic clarifications for the moment, because the diagnostic imaging division sometimes uses some of these terms in a slightly different way.

And my assumption is that your comments are relating to using the surrogate for licensure of a therapeutic. We tend to also in our division talk about how we validate an imaging product for approval perhaps being licensed to be used in this context. So you're talking about the former, using it in a true surrogate sense, reasonably?

DR. HUGHES: Absolutely, yes.

CHAIRPERSON KAWAS: Thank you very much.

I want to thank all of our excellent speakers this morning who were very informative and stayed to time, which is why they get a whole hour for lunch.

And I thank all of the Committee members for their excellent questions and their attention.

The members of the Committee will have a

special place in the dining area where they can eat. And we will plan on reconvening this meeting at 1:45, in an hour.

(Whereupon, the meeting was adjourned at 12:40 p.m., to reconvene this same day at 1:49 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

CHAIRPERSON KAWAS: The meeting of the FDA's Advisory Committee for Peripheral and Central Nervous System Drugs is now reconvened.

We'll be beginning with the open public hearing.

I'd like to remind all the speakers during this portion of the program, that in the interest of the fairness they address any current or previous financial involvement with any firm whose product they wish to comment upon.

And our first speaker for the public hearing is Dr. Eric Reiman, University of Arizona, Good Samaritan PET Center.

DR. REIMAN: Well thank you very much.

I wanted to offer some personal recommendations on the use of brain imaging in Phase III clinical trials from the perspective of a brain imaging researcher whose interested in using these techniques in the evaluation of drugs for the treatment and prevention of Alzheimer's dementia.

I have no current financial arrangements

with industry. I have served as a consultant to Pfizer, Elan, GlaxoSmithKline, Solvay and Meinse with regard to the role of these imaging techniques and FDG PET in particular in early detection and tracking.

In my opinion, brain imaging techniques should provide ancillary measures of clinical efficacy in Phase III clinical trials, and information about disease modification in these trials.

To date I believe the published data support the use of volumetric MRI and FDG PET in the prediction of a drug's clinical benefit in that they are reasonably likely to predict a clinical benefit.

I also believe that they are reasonably likely to determine the extent to which a drug's benefit is related to disease modification.

As you have heard, published studies for both of these modalities have suggested improved statistical power over traditional outcome measures and other neuropsychological test measurements, for that matter.

And while I think there is reason to support its use for disease modification if that

support is provided, then those studies when these imaging techniques are embedded in clinical trials, we will then have the foundation to validate this surrogate markers.

And of primary interest to our group is that the validation of these surrogate markers is absolutely critical for their use in the efficient discovery of prevention therapies. Not only secondary prevention therapies in patients with mild cognitive impairment, but primary prevention therapies in cognitively normal persons at risk for the disorder.

To date it is very hard, in some cases impossible, to test the efficacy of a promising primary prevention therapy. It is impossible, for instance, to study a hormone replacement therapy if it was presumed to be safe soon after menopause and determine the risk of developing mild cognitive impairment or Alzheimer's Disease. And I think these techniques have special promise in that regard.

Of the imaging techniques that are out there, I believe that volumetric MRI and FDG PET are the imaging modalities of choice for these trials. In

particular, as you've heard, MRI measurements of hippocampal entorhinal cortex and whole brain volume, and FDG PET measurements of posterior cingulate parietal, temporal and pre-frontal glucose metabolism, published studies have supported their potential role in predicting a drug's clinical benefits and determining the extent to which the changes reflect disease modification.

As you've heard, for each of these measurements cross-sectional studies have shown a correlation with dementia severity, studies for most of these measurements have shown prediction of subsequent clinical decline and also prediction of the histopathological diagnosis of Alzheimer's dementia.

And there are longitudinal data for each of these measurements now that indicate that the changes are progressive, provide data for preliminary power estimates and suggest greater statistical power than traditional outcome measurements. As you've also heard, these declines precede the onset of dementia. For the MRI measurements, we have good data showing the parallel on decline with memory concern prior to

dementia. For the FDG PET measurements we have good data showing these declines precede the onset of any cognitive impairment in carriers of a common Alzheimer's susceptibility gene that one out of four of us have, providing a great promise in the study of prevention therapies once these markers are better validated. And to validate these markers, we have to have these imaging techniques embedded in clinical trials.

I strongly believe that two imaging modalities are better than one. That the use of both volumetric MRI and FDG PET in Phase III clinical trials can largely address most of the surreptitious effects that have been discussed to a large extent in addition to randomized start or withdrawal trials. These complimentary measures of brain function and brain structure together can provide converging evidence in support of a drug's therapeutic effects. And when used together are very likely to provide information about outcome and disease modification.

Together they increase the certainty that the effects would predict outcome and reflect disease

modification. This is less relevant for Phase III clinical trials, but for proof of concept studies, one could also imagine an unlikely confounding effect on one imaging modality that minimizes one's ability to detect disease modification effect. If, for instance, in the unlikely effect that the removal of plaques shrinks the brain, one would still have another measure for proof of concept studies. Less relevant for this issue, but there are numerous benefits to the use of both measurements in increasing our certainty that our findings will be relevant to predicting clinical outcome and disease modification.

I believe that the combination of these techniques will provide the best foundation for the development of these likely surrogate markers in providing information about at least one's a valid surrogate marker in the future. And I believe that they provide the best foundation for establishing their relative roles in the efficient discovery of prevention therapies.

I believe that the use of the combination of these techniques for the reasons I've described,

that the additional cost is more than justified. Both imaging techniques are widely available. And I believe that the logistical challenges can be readily addressed in performing both studies using both modalities in these subjects.

So in conclusion I'd like to suggest that volumetric MRI and FDG PET should provide ancillary measures of efficacy in Phase III clinical trials, that they are likely to predict outcome, that there's reason to give industry the incentive to get a label for a disease modifying effect because of that reasonably likely criterion. And that once that's done and these studies are used, we'll have several additional long term benefits of their use.

I believe, as I've mentioned, that the combination of MRI and PET is justified at this time and could help address some of those lingering uncertainties. And I believe that the long term benefits of using these techniques in Phase III clinical trials are extremely important, the further validation of these surrogate markers and the development of a way to discover prevention therapies,

including primary prevention therapies without losing a generation along the way.

Thank you.

CHAIRPERSON KAWAS: Thank you, Dr. Reiman.

Our next speaker is Dr. Mary Pendergast from Elan Pharmaceutical Management Corporation.

DR. PENDERGAST: Good afternoon, and thank you for allowing me the opportunity to present to you this afternoon.

I am Mary Pendergast, Executive Vice President of Elan Corporation, Elan Pharmaceutical Management Corporation, the holding company.

Part of Elan develops and sells genetic and other tests for Alzheimer's Disease. And another part of Elan is working to develop therapeutics for Alzheimer's Disease. Elan does not have an interest in any brain imaging modality or technology.

In my written statement to the Advisory Committee I explained why a surrogate marker does not need to be validated before it is used in drug development as a primary endpoint. Rather, any surrogate marker that is reasonably likely to predict

clinical benefit can be used to approve a therapy so long as trials studying clinical endpoints are carried out later. On that point I think there is an agreement between the agency and myself.

I also think that if there is a surrogate endpoint that is reasonably likely to predict clinical endpoint, the FDA must permit its use, even when the agency might prefer to wait for validation of the surrogate endpoint or the agency might wish that there were trials using well-defined clinical endpoints. The agency may not agree with that. It probably does not want to have its discretion curtailed, but I think that that interpretation is the only way to give meaning to the congressional directive that FDA must facilitate the development of fast track drugs.

In any event, any surrogate marker that is reasonably likely to predict clinical benefit, I would argue in that circumstance the FDA should and would want to use the surrogate marker because Alzheimer's Disease is a serious public health problem.

If you look at the slide, I'm sure you've all seen the billboards on the buses around Washington

and other cities. 40 million persons infected with AIDS, zero million cured.

The same is true for Alzheimer's Disease. 15 million infected are infected, zero million are cured. And if you look at this slide, you'll see that in the United States there are four times as many people that have Alzheimer's Disease than have AIDS in this country.

Surrogate markers have made it possible to develop disease modifying therapies for HIV infection. There are no disease modifying therapies for Alzheimer's, and I think one of the reasons why is because we haven't started using surrogate markers yet for drug development in Alzheimer's.

We need to use surrogate markers to develop drugs for Alzheimer's because by the time the patients have full-blown Alzheimer's Disease, or even the inappropriately named Mild Cognitive Impairment and they are showing symptoms, they have lost a significant amount of their power. They have suffered probably irreversible neuropathology. While drugs that might have symptomatic effects might be

relatively straightforward to study using clinical endpoints, that may not be the case for disease modifying drugs. Based on animal studies, disease modifying drugs may not show immediate symptomatic relief, but rather by attacking the underlying pathological cascade, they might slow the rate of neurodegeneration.

Given the variable course of Alzheimer's Disease, trials showing a change in the slope over time, even in MCI or AD patients, will be large and long and a surrogate endpoint might tell us more quickly whether the treatment is working or failing. We will be able to learn whether the drug is having an impact before the trial participant dies or becomes yet more demented.

Perhaps more importantly, surrogate markers will permit us to study drugs at earlier phases of the neurodegeneration, before the neuropathology becomes severe enough to be manifested by clinical signs and symptoms. We want to be able to study and ultimately treat patients in that 15-year period when the neurodegeneration is taking place, but

the symptoms are not yet troublesome.

We should also remember that the clinical endpoints currently used are somewhat crude. For example, ADAS Cog has a huge standard deviation. Ten years from now we will probably think of our current clinical endpoints the same way we now think of the earliest definition of Acquired Immune Deficiency Syndrome, which was a rigid definition based on clinical symptoms, that turned out to miss many patients with HIV infection who needed therapy.

As with HIV, in Alzheimer's Disease the more valuable endpoints will probably be the surrogates.

In summary, there are several types of brain imaging modalities: MRI, MRS, PET. If they are well established or validated as Dr. Hughes has described, they can be used as primary endpoints for traditional approval. But even if they are not well established, even if they are not validated, they can still be used to support approval of fast track drugs with confirmatory trials to follow.

I'd like to point out that there was five

years between the time the FDA first approved a drug for HIV infection based on HIV PCR and the time HIV PCR was validated as a surrogate endpoint. That's five years of patients that received treatment, that's five years that patients got a treatment that could keep them alive long enough for the next therapy to come down the pike.

There are examples other than cardiology that can be used with respect to surrogate markers. I've mentioned HIV. There are other diseases as well. In my written statement I point out the analogies that could be made between rheumatoid arthritis and Alzheimer's Disease, diseases where you have both endpoints based on the signs and symptoms of the disease and measurements based on the structural damage that the disease causes.

I urge you to think of the drug development landscape broadly and find, as with HIV, rheumatoid arthritis, cancer and other diseases that surrogate markers have an essential role to play.

The questions you need to ask yourselves are not difficult. They are a question of risk. Is

slowing the rate of cerebral atrophy reasonably likely to correlate with clinical benefit? Is slowing the rate of accumulated tangles and plaques reasonably likely to predict clinical benefit? Is slowing the rate of decay from normal metabolism to hypometabolism reasonably likely to benefit the patient? I mean, ask yourselves the question: If this was your brain, would you want it to shrink, get plaques and tangles and become hypometabolic? I wouldn't.

I think one more point I would like to make is that this is a question of risk. And one of the ways risk comes up is in a question with respect to safety. Because it is definitely true that if you approve a drug on the end of a couple of Phase II trials using a surrogate marker, which has been done many times before by the agency, you will not have the same large safety database that you otherwise would have had. But both the agency and its accelerated approval regs in 1992, and Congress when it passed the Fast Track legislation in 1997 had solutions to that problem. The solutions are several-fold.

First, they require the companies to

continue to study the drug out to their clinical endpoint and out to a large safety database. And those trials can be compelled and they have been compelled.

Second, the agency can demand additional safety reporting and monitoring by the drug company during this period when the drug is approved on a surrogate and when the final clinical trials are finished.

Third, the agency can restrict the distribution of the drug to practitioners with certain academic degrees, to tertiary medical centers, to whatever they feel they need to do for the safe use of the drug. And they can limit and in fact completely exclude the ability of the companies to advertise about the drug.

And finally, Congress recognized that the agency will make mistakes with surrogate markers. It's inevitable. And so there are very easy ways of getting these drugs off the market.

When we first invented this system when I was at the agency, we called it "easy on/easy off."

So, with that, I'll answer any questions

you might have.

Thank you.

CHAIRPERSON KAWAS: Thank you.

Is there anybody else in the room who would like to make a comment during this time. This is the last chance anybody in the audience or otherwise may have. Okay.

This concludes the public hearing portion of this meeting, which takes us to what I consider to be the hard part, although it seems like a lot of people in the room don't think it's going to be nearly as hard as I do; the discussion of the issues presented by the FDA.

So, I'm going to open the floor for the discussion in a moment. I want to remind everybody that in addition to discussing the presentations that we've heard in general and in specific, that we also were given several questions that we're supposed to be focusing our thoughts on. And those questions have been provided to all of the Committee members, the first one of which is: How is the surrogate imaging modality best validated?

So, with that I open the floor for the discussion of the Committee on how is the surrogate imaging modality best validated?

DR. SORENSEN: I was wondering if I could ask a question of, I think it was Dr. Hughes. As I was listening to your presentation and thinking about it over lunch, I wondered if any surrogate endpoints could ever be considered valid? I mean, you showed some data of where cholesterol failed, and I guess there was some discussion, Dr. Temple mentioned, about hypertension before the definitive studies were in. And I just wonder how -- and yet we have some that have been used by the agency. I'm trying to figure out how we get over that -- we make that decision.

It seems like you described that there were a lot of ways to show that a surrogate didn't work if you have single trials that show that there is a discrepancy. Other than meta-analysis, it seems like that's probably the way. And even those, you had exceptions to all of your meta-analysis.

Is there a time when somebody can say, you know, the Cochran report is out, we're done, or how do

you actually kind of make that decision?

DR. HUGHES: I think you have to put validity in the context of risks of using a surrogate. And I think you can reasonably say based upon both clinical trial data and epidemiological information that anti-hypertensive effects, cholesterol lowering effects, effects on viral load, possibly effects on CD4 count in HIV are good surrogates. And that if you base decisions about the effectiveness of a drug on the markers based upon past experience, you're very unlikely to make an error.

DR. SORENSEN: And so is that a database of 10,000 patients, of 50,000 patients or what?

DR. HUGHES: I would say in each of those areas you're probably talking about a database of 20 or 25 large, randomized trials and in each of those cases I think there's epidemiological evidence after the sort of validation process is being conducted to show that you were right.

In other words, if you take HIV as an example, the effect on the marker is dramatic and the effect on clinical outcomes has been dramatic and you

can see it in surveillance data in the U.S.

DR. SORENSEN: Yes.

DR. HUGHES: So I think in those contexts for the types of interventions that were being studied or evaluated, the risk of inappropriate approval is probably minimal.

Now, having said that and if I take HIV as an example, that's been demonstrated for antiviral drugs within certain classes. That doesn't mean that those same markers would work well for, say, immune based therapies.

And I don't think, for instance, the FDA would necessarily advocate the use of those markers for immune based therapies.

DR. SORENSEN: Sure.

DR. HUGHES: So I think in some areas I would consider the markers have been validated in the sense that the risks of using those markers for certain classes of interventions has been minimized.

DR. SORENSEN: Well, so then just to finish my comment, I guess I wasn't counting exactly how many studies were presented today, but I don't

think there were 25 randomized trials of this. At least it doesn't look like we've got a validated surrogate endpoint for brain imaging in Alzheimer's as of yet.

CHAIRPERSON KAWAS: I think that probably a large number of people in the room would agree with that statement.

I have a question for any of the invited imagers or anyone else.

We've talked about human studies today and we've seen some interesting human data, although not 25. Has there been any work done with animal models to show that these interventions and these measurements may be relevant? And if so, can someone share some of that with us?

DR. De CARLI: It's not my own data, but there's mouse models showing that there is brain atrophy accompanying the progression of the disease. They were abstracts presented at Stockholm. I don't know that they've become full papers, but I think they will be shortly. But there is some preliminary animal data that shows atrophy associated with the

progression of the disease.

CHAIRPERSON KAWAS: And is there any data that shows that interventions, for example the vaccination mice, has anyone imaged their hippocampal volume or --

DR. De CARLI: I haven't seen that data yet.

CHAIRPERSON KAWAS: Or their NAA?

DR. De CARLI: I haven't seen it.

MS. ROBERTS: Having applied it to therapeutic modality, but rather than using FDG PET, we used FDG autoradiography and PDAPP transgenic mice and found that they had a preferential and progressive decline in posterior cingulate glucose metabolism, the one brain region that is homologous to that in the humans suggesting that dysfunctional brain imaging measure might provide a way to track disease progression in the animals and screen candidate treatments. But that needs to be extended to other mouse strains and confirmed in other studies.

CHAIRPERSON KAWAS: Dr. Small?

DR. SMALL: We've also done studies with

transgenic Alzheimer's mice with FDDNP with autoradiography and found increased cortical signal compared to control mice.

One of the challenges with micro-PET is that the mouse head tends to be a bit too small to pick up the signal. So if we could get some good transgenic rats, we might be able to get a little bit farther, they tend have bigger brains.

DR. FOX: With your permission, I'd like just to make a comment on the question from Dr. Sorensen. Would that be all right?

The example given of hypertension is perhaps the -- in terms of number of patients, perhaps the most validated surrogate. I think it would be worthwhile to think about some of the hypothetical possibilities that we've drawn up about brain imaging as a surrogate, for which I accept there are lot of possibilities where you could alter the surrogate without altering the outcome.

But if we take hypertension, venesection, massive venesection would probably reduce your blood pressure but might not alter the outcome in the way

you would hope.

So I think it's always possible to put out some possible hypothetical example of where the surrogate would fail, and it's all down to understanding or trying to understand the pathological cascade. So I think, you know, a massive venesection might not be a good effect on clinical outcome, but would effect blood pressure.

CHAIRPERSON KAWAS: Thank you.

Let me try and summarize and then everyone can tell me how I mis-summarized.

I mean, it sounds to me from the discussion that we've heard so far and the invited speakers who showed us data that overall the general sentiment is the best way to validate a surrogate marker would be in human studies by combining multiple studies. And I seem to have heard a lot of calls for putting imaging into ongoing clinical trials in order to be able to do that.

Is the answer potentially to our question that the best way to validate a marker is to continue on doing human studies as opposed to trying any other

alternative approach?

Dr. Grundman?

DR. GRUNDMAN: Yes, I agree. I think we need to do human studies. But I think what we really need to do is a large multi-center type study where we look at serial PET and MRI in conjunction with the cognitive and clinical outcomes and see how they predict the clinical outcomes in a really rigorous prospective fashion. I think that would give a lot more credence and credibility to the field in terms of using them as a marker.

CHAIRPERSON KAWAS: Which is basically how the clinical trials for the most part are being done now, at multiple sites. So superimposing it on the trials would be a strategy as long as it met those requirements?

DR. GRUNDMAN: Yes. I mean, the other problem, you know, in terms of validating a marker, in terms of requiring that a drug actually modify the surrogate and modify the disease outcome is that, you know, this may not be possible. We don't have drugs right now that can do that. So I would say in the

absence of a drug effect that we can definitively state has an effect on the outcome, even an observational study, a large observational study that could make the correlations between the PET and the MRI and the clinical outcomes would be a reasonable approach right now.

CHAIRPERSON KAWAS: I have another question, actually, for our invited speakers.

Dr. Fogel, did you have --

DR. FOGEL: Yes. It seems that I agree with Dr. Sorensen that we don't really have a specific surrogate at the present time that looks like we can use. And I was wondering if maybe Dr. Hughes might be able to address it.

Even though we don't have just one singular surrogate, I guess I'm wondering in terms of using a, for lack of a better word, a composite surrogate? We've heard of a number of candidate surrogates that one might be able to use. But I'm wondering from a statistical standpoint, and I guess from a study design standpoint, would it be better to -- or how hard would it be to combine some of these

surrogates to make a composite surrogate, if you will, weighted or unweighted and then use that composite surrogate to be able to tell whether or not that goes towards the clinical outcome? Because it doesn't seem from looking at all the data that any one will do it. But if we combine two or three and weighted it, or however one wanted to do it, whether or not that might be a useful approach to using surrogacy for clinical outcome.

DR. HUGHES: I think that's an excellent point. I think the way that you would validate a composite would be exactly the same as you would validate any individual measure. So I don't think it changes the validation process. And you've got the same problems with lack of information.

DR. FOGEL: Although if you have a number of different, you know each one has a certain percentage to correlate, for lack of a better word, with the clinical outcome. And I guess I'm just wondering if each one has that certain small percentage, the intersection of all three might be more specific than either one together. And I think

the crux of the problem here is that it's not specific. These things can go awry in many different ways.

I believe somebody talked about using amphetamines to increase glucose uptake as opposed to just being due to Alzheimer's. And I guess the question is if we meet at the intersection of 3 or 4, or however many people eventually decide might be a good thing, that that intersection point might be the goal that we want to reach rather than any individual one.

DR. HUGHES: No, I think you're quite right that you could create a composite which would be much more specific. And I guess the way that you would start going about that would be in, for instance, natural history studies to create a prognostic indicator based upon several measures which would be a better predictor of ultimate outcome. And then having done that, validate it in much the same way as you would validate an individual measure within a clinical trial.

DR. FOGEL: I mean, there's precedence in

the congestive heart failure world about using composite endpoints to look at the efficacy of drugs. And I'm just wondering whether or not a similar framework might be useful in Alzheimer's Disease as well. So it isn't like there's no precedent for it; there is.

CHAIRPERSON KAWAS: Dr. Provenzale, and then Dr. Wolf.

DR. PROVENZALE: Thank you.

I think one of the fundamental issues we're trying to grapple with here is how to make the jump from a prognostic marker, even a very good prognostic marker, to a surrogate marker. And I think part of the discussion has gone along the lines of, well, maybe if we combined a number of prognostic markers, does that make a surrogate marker.

And I'd like to, you know, ask the question of the group what is it that is presently or what do people feel is presently lacking that would make the difference, that would push us over the hump, as I think Craig was getting at? And I don't think the answer is simply just tacking on more and more

prognostic markers. But as Dr. Grundman kind of pointed out, the problem is that we don't have a drug that effectively treats this disease, so how do we somehow pull surrogacy out of this? I'm sure it's possible, but I think that's where we're stuck.

CHAIRPERSON KAWAS: Do you feel the need to respond?

DR. GRUNDMAN: No. I was just going to say, basically that's the problem, we're in a catch-22. You're saying we can't have a surrogate unless we have an effective drug. So once we have an effective drug, we won't need a surrogate anymore.

DR. PROVENZALE: You have to pull yourself up by your bootstraps, and how do we do that?

DR. SORENSEN: And you need 25 studies.

CHAIRPERSON KAWAS: Dr. Katz?

DR. KATZ: Well, I guess it matters what you mean by all of this. Again, there are several concepts. One is how do you validate a surrogate or the question I probably would be more interested in hearing responses from the Committee on is whether or not anybody thinks that any of the surrogates proposed

today, or any other surrogates for that matter, candidate surrogates have actually been validated in the ways that Dr. Hughes talked about, and a I talked a little bit about? I think I know the answer to that question, but I still think it would be useful to hear people talk about that.

As far as, you know, sort of this catch-22, of course if we don't have a drug that has an effect even on the surrogate, let alone whether we know it has a clinical effect, of course we couldn't approve such a drug. But as you've heard from various people, we do have a standard, an alternative standard for the approval of drugs, so called fast track drugs, which we can impose.

Now, Mary Pendergast suggests that we must impose it. I'm not sure. I'm not sure that there's that much difference, quite frankly, in our views.

But nonetheless, there is a standard. And that standard says reasonably likely to predict. So at some point if you think that no candidate surrogates actually have been validated, we have to discuss whether or not anybody thinks that in the absence of

any clinical finding for a particular drug, whether or not an effect on one of these potential candidate surrogates is reasonably likely to predict. And then if we pick a surrogate, let's say MRS or NAA or whichever one you might pick -- you might pick none of course -- but if you picked one, we could use that as the standard to test the next drug that comes down the pike. And it either will have an effect on the surrogate or it won't.

So I don't think we have -- I think you have to worry about having a drug that does this when you talk about validating a surrogate. But if you want to impose the standard of reasonably likely which permits the approval without expressly and explicitly without validation, then you'd pick one, and we'd use it.

So, I don't think we have to worry so much about the so-called catch-22 or the absence so far of such a drug.

CHAIRPERSON KAWAS: I'd like to give Dr. Wolf a chance to speak. But then would it be helpful to you, Dr. Katz, if we sort went around the table and

let each person express whether or not they think any of the images that they've seen today have been validated for use as a surrogate and if so, which ones?

DR. KATZ: Yes.

CHAIRPERSON KAWAS: Okay. Dr. Wolf?

DR. WOLF: Well, my question will address this partially. Because one of the problems we have with many of the imaging modalities we have seen is that the techniques and procedures that were used for many of them were quite different. And although we have a number of studies that use, for example, MRI, because they use different protocols they are not strictly comparable.

So one of the problems we need in developing the prospective studies is to have a uniform, well thought out protocol so that we can compare studies across multi-centers. And right now we have a number of common studies that use the technology, but which are done in a different manner and therefore, are not necessarily strictly comparable.

So, for example, we have seen some data where study X got positive results, study Y got negative results. And they're probably both done correctly. But because they were doing things in a slightly different manner, their results came out differently.

So this is one of the problems that we have to face. I mean, what is the time resolution, what is the spatial resolution, what's the degree of localization; what are a lot of these parameters we use in imaging modalities and how comparable are they from one side to another.

CHAIRPERSON KAWAS: Thank you, Dr. Wolf.

I think that Dr. Katz actually gave us two separate questions, and I'd like to sort of sort them out as we go around the table and let everybody give their opinion.

So the first question is have any of these markers been validated for use as a surrogate, at least on the level that the individuals believe it should be. The second question, which we will take up later, is the reasonably likely possibility that any

of these markers may be useful, and which of those markers have met that level of standard.

So, to begin with, can we start with the right side of the table and we will give everyone an opportunity to answer the question whether or not they think any of these modalities have been validated as a surrogate marker in the disease of Alzheimer's.

Dr. Provenzale, I think --

DR. PROVENZALE: It's my opinion that none of these have been validated at present as a surrogate marker.

DR. FOGEL: No, I don't think any one of them have been validated either, although I would like to at some point get back to the concept of whether or not by a meta-analysis one could use a composite, and the data may even be there it would be valid but we don't know it because the analysis hasn't been done, a combination of various surrogates.

DR. VAN BELLE: I don't think any evidence has been presented yet that would convince me. And especially because I think as we heard this morning, and I think this makes sense, you would only be able

to establish effectiveness if you had a series of randomized clinical trials.

On that point, given the claim for improved power, it should be relatively easy to incorporate these candidate surrogate endpoints into clinical trials because presumably they're going to have bigger power than some of the other clinical endpoints.

CHAIRPERSON KAWAS: Good point.

Dr. Penn?

DR. PENN: I think for all effective purposes the MRI quantitative measurements of atrophy have been validated for being quantitative measures of atrophy; just that. And that it is reasonable to use those as a surrogate. And that it does, in fact, measure disease and that within 10 or 15 years we will be looking at the disease that way rather than looking at clinical manifestations. And it's going to be a painful thing to go through this transition, but I think that it's very likely that it'll happen in the same way it's happening in MS now for using MRI to show the disease itself.

CHAIRPERSON KAWAS: Okay. I think my opinion is that none of the markers to date have been adequately validated for use as surrogate in studies of AD.

DR. GRUNDMAN: I would very much like to have a valid surrogate marker, but I agree with you. I don't think that we actually have one. I think, you know, if you look say just at the MR measures and you wanted to have some sort of standard outcome in the clinical trial that we could use as a standardized measure, I wouldn't know right now whether or not we should look at the hippocampus, or we should look at whole brain, or look at gray matter or look at some other measure. What's the best measure of clinical progression as it relates to a standard clinical outcome measure that we would look at in the clinical trial. And I don't know the answer to that. And that's why I think we need to do those sorts of prospective large scale studies to make those correlations in a sort of definitive way and figure out what the best measures of brain atrophy actually are.

And I can say a similar thing about PET.

You know, there were different regions of the PET scan that showed decreased metabolic rates and, you know, posterior cingulate frontal, temporal, you know. I'm not sure which, was it the whole scan, is it part of the scan? What particular segment would you be looking that? We saw some measures that correlated on the left side of the brain that correlated better with the MSSE than others.

I think at this point we just don't know which measures even of the PET scan are the best or most closely associated with the outcome measures that we're interested in in a clinical trial.

CHAIRPERSON KAWAS: Dr. Wolinsky?

DR. WOLINSKY: So the short answer is no, these are not validated surrogates. The longer answer, which I feel compelled to give, is that there is very intriguing data that's been presented here and outside of the room that says that quantitative image analysis and functional imaging of various types is our only portal to the pathology of brain disease. And I'm not at all comfortable that any of our current "clinical outcomes" are more reliable than

these portals will be in the long run. And, in fact, I'm very discouraged that at least some of the things that I deal with are not telling us a good picture of what goes on.

The issue, though, is a little bit different and comes back to the first answer, which is, no, these are not proven surrogates, they can't be in the definitions that we've been given to work under. And maybe after we get around the table if there's time for other things, we could maybe think about more novel ways to use these kinds of critical tools in trial design that might be more useful in dissecting what happens in trials. But that's a longer statement.

DR. CHIU: Up to now, the clinical imaging -- because I'm stating from an MI point of view, and we read the MI and then we do a profusion, we do a PET scan. And we do see it. Actually, today it's listening to lectures. We lack of understanding what MI can do. So we should have more education in this regard. Because MI up to now, it's throughout whole United States. It's only you have 1.5 test results.

We don't have to go any -- you know -- good scanner. And we can perform it, we can really see the cortical atrophy. We do see a hippocampus abnormality, even though not specific for AD. But we do see the changes.

So I believe in the MI image and PET and the fusion image.

CHAIRPERSON KAWAS: So are you saying they have been validated or they just have potential promise, need further study?

DR. CHIU: From our center's point of view, the clinicians who believe in that, we continue to --

CHAIRPERSON KAWAS: So you believe that they have been adequately validated as surrogate markers for use in drug trials?

DR. CHIU: That's correct.

CHAIRPERSON KAWAS: Dr. Ramsey?

DR. RAMSEY: I would agree with Drs. Grundman and Wolinsky, and for the same reasons that they gave, but I'll withhold my opinion from PET since I don't know enough about it.

I hesitate a little bit because I think

there are a lot of people behind me who are probably a lot smarter than I am who think that they are valid markers. So that bothers me a little bit. But even as they were presenting their information, I kept feeling, as Dr. Chiu did, that I want to see more images, can I see a little more about that. What about the T2 weighted images, are there a lot of hyper increased signal intensity areas, is that what's really depressing the NAA? What else is going on? Do they have seizures, is that what's really affecting the temporal lobe?

So all of those concerns. And maybe it's just because in this short period of time we can't present all the data that's out there, but I feel that from what we saw and what's available, it hasn't really validated as an acceptable surrogate.

CHAIRPERSON KAWAS: Dr. Beam?

DR. BEAM: My short answer is simply that I don't know. I wish I could say yes or no at this point in time, but given the data that I've seen in the short period that we've been here, I just can't make this determination right now.

I would like to abstain from the question on the basis of simple ignorance. I would like to have more discussion about this and perhaps a longer presentation of the existing data might lead us to a different conclusion in the future.

DR. WOLF: My concern is that I'm not sure how valid and how meaningful the clinical data are and to what extent they are definitive and they are truly a gold standard.

If we go to the basis that the current clinical procedures are an absolute gold standard, then I'm not sure the imaging modalities are yet proven to be equivalent. If on the other side, we have concern that the clinical measurements are also fraught with a lot of uncertainty, then probably the imaging modalities are close in uncertainty. And under the circumstances I'm not quite sure to say yes we can discard them. I think we need to consider them. I think we need to for each drug we need to consider the weight of the evidence.

And if imaging modalities provide enough supporting evidence that reinforces and supports some

of the clinical data, then they must be considered as part of the package.

As single systems because they are not part of the traditional standards, I don't think we want to go for that. But at the same time, we need to continue looking at them because, like Dr. Wolinsky said, I don't think we have a good measurement at the present with the clinical outcome.

CHAIRPERSON KAWAS: Okay. Dr. Sorensen?

DR. SORENSEN: Yes, it's also my opinion that the standard for validation of surrogate endpoint has not been met by any of the data we've seen so far.

CHAIRPERSON KAWAS: Thank you.

Dr. Kim?

DR. KIM: From the data presented today and some of the literatures that are available, I see changes but I'm yet to make the connection between the changes that we see here on the data and what we see it as a connection between that change and the AD. So I'm still yet to be convinced with that.

CHAIRPERSON KAWAS: Thank you.

Did you want to --

DR. VAN BELLE: No. I think we got a pretty clear view of whether or not people in general think that any of these have been validated in the sense that we've been talking about. And that's very helpful.

If I can move to the next question, which is, given that the consensus is that none of these have been validated, the question then arises whether or not we should rely on the drug's effect on a surrogate -- I'll leave for the moment which one or which ones -- whether we should rely on the effect on the surrogate solely in the absence of clinical changes to support the approval of a treatment for Alzheimer's Disease.

As I pointed out earlier, and as Mary Pendergast pointed out, we have language both in the regulations and in the statute, in the Act, the law, that say that we at the very least can approve a drug on the basis of an effect on a surrogate that is reasonably likely to predict a clinical benefit or to represent the clinical benefit like, for example, progression.

I know you've heard the term all morning and afternoon "reasonably likely." And, of course, I can't give you a lot of guidance as to what that means, although the language in the regulations talk about epidemiologic, pathophysiologic or other sorts of evidence. That's not very helpful. But the question now given that you believe that no surrogate is validated is should we rely on the effect on a surrogate in the absence of a clinical change at this point, at this time, to approve a drug for Alzheimer's Disease, which if you say yes, you have had to have concluded that it was reasonably likely to predict or to represent a clinical change. And if you do say yes, I'd be very interested to know how you've come to that decision.

But that's the question: Can we approve a drug on the basis of an effect on an unvalidated surrogate in the absence of a showing of a clinical effect?

And again, I would ask you when you think about that to take into consideration Mary's point, which was that there is a belief, anyway, that a drug

that would have an effect on progression may not have an effect that can be seen clinically very early. Right now the symptomatic treatments can show effects in 6 weeks, 3 months, 6 months certainly. But there is a question as to whether or not a drug that has an effect on the underlying progression as represented, perhaps, by a surrogate will show clinical benefit early.

So, anyway, that's the question we critically need you to discuss.

CHAIRPERSON KAWAS: Dr. Van Belle?

DR. VAN BELLE: Apropos to that point, I've been pondering a graph that Dr. De Carli showed from the Framingham study where he has data relating the brain volume from age 30 to age 95, basically I don't know whether you remember that graph or not. But it's clear that there was a very steady progression of decline in brain volume from age 30 on.

I wonder if he had superimposed on that, say, the MMSE that might have been estimated at the same time, whether that would have shown a decline as well or whether that's pretty standard?

The point I'm trying to make is that I'm not convinced yet that changes in brain volume are necessarily associated with changes in cognition. And I think that's really a prerequisite for dealing with one of these imaging surrogates as a possible modality for a clinical endpoint.

CHAIRPERSON KAWAS: Thank you.

Dr. Sorensen?

DR. SORENSEN: Yes. I'd like to respond to Dr. Katz.

I think -- I was hoping you were trying to set up a kind of a straw man by saying is there any chance that one would be happy with an agent that didn't have a clinical benefit but did have an MRI benefit. And certainly if one had the option to have both a clinical benefit and an imaging benefit, you would certainly take that option.

And so my initial response was of course not, that wouldn't be feasible. But then I got to thinking about kind of potential scenarios where the answer -- where I could try to find a way to say yes to that. And the best analogy I could come up with in

the few moments is the coded stems that we've seen such dramatic results from angiographically, and yet it might take, you know, years to provide that their clinical outcome had some meaningful benefit in, say, survival of patients.

And so I guess I can imagine an scenario where someone might have a complete cessation of atrophy that had been, you know, documented before they were on the drug and then they stopped. And that the MMSEs or the ADAS Cog tests were trending towards a positive impact, but they hadn't actually reached a positive impact. And so you'd have to say there is no evidence that statistically that there was a clinical benefit.

You know, would I want to at that point say to patients that this drug couldn't be approved? I think at that point I'd probably go squishy and say "Well, let's look at the safety profile, let's look at some of the other mitigating factors." Because we don't really have good tools to understand what's going on in the brain. And if this one marker, whether it was the ADAS Cog score that we knows it has some --

some challenges or whether it was an imaging score showed a lot of benefits. Even if the other ones didn't. I would hate to close the door on that.

So, I think that the challenge of prospectively defining what that is, what that reasonableness is, I think is very hard. But to say there's no scenario at all under which I could come up a situation where I didn't have a clinical benefit, but I did have an imaging benefit, would I never allow that to lead to an approval? I don't think I'm ready to quite close that off completely.

DR. KATZ: I'm not -- if I can respond.

Yes, I'm not asking whether or not there is -- it's possible at some point or there is some imaging marker that at some point maybe, you know, I'm not asking if we should close the door forever and all time, or even if we should close it now. I'm asking should we open it now, really.

I'm saying right now do you think that there is an imaging modality, a surrogate marker if a drug was shown to affect it beneficially but have no clinical effect in a trial of some reasonable

duration; whether or not those sets of facts should allow us or should force us to approve a drug for Alzheimer's Disease now.

DR. SORENSEN: Okay. So you're not closing it quite -- sorry. I'll just finish the point if that's all right.

I see. I thought you were going to try to take -- peel away from one extreme down to sort of the reasonableness issue. And I guess I would still say that those markers that could lead to success that would I think be compelling evidence that a drug might have benefit, could even be the ones we've seen today as unvalidated as they were. If somebody came to me with a set of data that was large and the imaging was done well, and they had a logical scientific argument, and they just barely missed by their clinical performance scores, I'd certainly be very tempted to seriously consider that, and would have to weigh in on other aspects. I'd have to look carefully at it and I wouldn't want to close the door to that.

DR. WOLF: I would like to support what Dr. Sorensen just said and expand it a little bit.

If the imaging modality shows definite positive results and the clinical outcome is not deteriorating, does not show a significant deterioration, then that drug may be considered seriously.

If on the other side there is a positive imaging outcome but clinically the patients continue deteriorating, then obviously the imaging modality cannot be weighed over the clinical arena. But the question is is when we have the borderline situation, whether there's no significant deterioration from the clinical point of view.

One of the things we don't know is what is a temporal relationship of what we measure. Are the imaging modalities giving us information that is earlier or later that with what we manifested clinically? And if in the case that the imaging modalities give information that manifests itself at an earlier stage, then shorter trials may reveal something that the clinical trial just have not caught up with.

So, again, it's something that needs to be

left open depending on the correlation between the clinical and the nonvalidated imaging modality.

CHAIRPERSON KAWAS: Well, I agree with both of the previous speakers, but actually I would like to contract the discussion back down a little bit. And if I understand Dr. Katz' question, I would like to respond.

As a person who works in Alzheimer's Disease and sees these patients and understands what the images look like in these patients, I really am absolutely -- I mean, I completely understand the correlation between the imaging and the patient's status. But what I don't find convincing, I think, and I've been trying to find all day, and I do think in the future we might have but I really believe very strongly we don't at the moment, is any evidence that makes me think that it is "reasonably likely" that altering these markers would necessarily have an effect on the disease.

I'm not convinced that we've seen anything here that couldn't just turn out to be hair color, and that aging would still go on and death would still

happen, and the hair would be black or the hippocampus would be bigger.

I think that superficially it sounds very tantalizing to assume that these things track very strongly with disease state, but I don't think I've been shown any evidence that makes me feel confident in saying it's reasonably likely that altering these parameters would have that effect.

I'd especially -- I want everybody around the table to try and give their thoughts. So, can I try going around again?

DR. PROVENZALE: Well, I was asked for my short answer to the last question, and I gave just a short answer.

But it's clear to me that probably one of the imaging techniques or a combination of imaging techniques that were presented today will prove very valuable in assessment of therapies for this disease. And so to get to the question when or under what circumstances should one feel comfortable relying on one of these imaging techniques as a reasonably likely to be successful surrogate marker, I would say that

there are probably a number of venues under which if well controlled prospective randomized trials in, let's say, that involved multiple sites all using the same techniques, the same pulse sequences, let's say, or the same PET imaging sequence, if they showed overwhelming evidence that one of these markers -- let's use hippocampal volume as an example.

Obviously, if we're talking about volumetrics, we could be talking about the whole temporal lobe, we could be talking about the whole brain, we could be just talking about small areas of the brain.

But, for instance, if there were a study in which a therapeutic agent was in a randomized controlled trial given subjects who were at high risk for developing AD but who at the beginning of the study all had normal hippocampal volumes, and if the study were executed properly and if a big difference were seen in the rates of change of decrease in hippocampal volumes, that would be to me fairly compelling evidence.

Obviously, as Dr. Wolf pointed out, we'd

have to take the clinical into consideration; that goes without saying. If the hippocampal volumes remain stable but the MMSE were deteriorating, that would be a different situation. But to me that would be very provocative and promising information.

Unfortunately, I think that we're stuck with a disease that progresses relatively slowly over time. And so we would not expect to see a dramatic change, a stabilization or improvement in MMSEs over months or a year or two. And so we have to, I think, more or less rely on markers such as this.

So although they're not validated, I think they offer as someone put it, that's our window into looking at this disease. I don't know which one it is. I think it's quite possible that a combination of the two, let's say thin section MR imaging for volumetric analysis with coregistered PET imaging or coregistered MR spectroscopy and PET imaging, or all three techniques together.

Although I don't think these are validated, I think we have to somehow figure out how we're going to use them to advance the field.

DR. FOGEL: I thank you.

Well, you know, because a surrogate by our definition means that we have to have an intervention that reliability predicts the clinical outcome, we obviously don't have in my opinion. So what all these great tests that we've been talking about falls into the realm of prognostic marker than surrogate. And so when we talk about reasonably likely to help the disease, we're really talking about reasonably likely using prognostic markers rather than surrogate markers.

And I guess I have a question for Dr. Katz, and that is we've heard a number of times already that these "surrogate markers" can be used in Phase II trials as drug picks to go on to further evaluation in Phase III trials. And we're essentially we're being asked is do we want to take this out of the realm of Phase II trials and enlarge this to Phase IV trials be unleashing it on the public. And so I guess I'm wondering how comfortable the FDA feels about taking stuff from Phase II to Phase IV on the basis of these prognostic markers?

DR. KATZ: Well, I think that's the question we're asking you folks. What we want to know, and I don't really -- that's post-marketing. But Phase II, Phase III, people have their own idiosyncratic definitions of what those mean.

My question to the Committee is do you think it's appropriate at this time to base an approval of a treatment for Alzheimer's Disease on the basis of a change on one of these candidate surrogate markers in the absence of any clinical change? I'm talking about the definitive trials on which approval would be based. So that's the question I'm asking.

DR. FOGEL: Under the most likely scenario?

DR. KATZ: Whether or not an effect on any of these surrogates. And by the way, for those who think that these surrogates are reasonably likely, it would be very useful to hear which ones do you think are.

But, yes, under the reasonably likely standard, whatever that means. Do you think it's reasonably likely that an effect on the surrogate in

the absence of a clinical finding is reasonably likely to predict a useful clinical outcome.

DR. FOGEL: And I guess in my opinion it falls, again, back to that we're dealing with do we have prognostic markers. Because those are the ones that would be reasonably likely to effect the disease. And we have a number of them that have been prognostic -- have been shown by data to be prognostic, meaning that they don't have an intervention but that have been shown that the marker itself has shown to change or to differentiate normal from disease state.

And I guess in my opinion from listening to all the data and reviewing some of the literature that we were given, I would vote for hippocampal volume and FDG. But, again, that would be under the reasonably likely scenario and not necessarily as a surrogate.

CHAIRPERSON KAWAS: But just so I make sure I understand your position, Dr. Fogel. If a study was brought forth today that showed that by giving somebody a compound you could alter their hippocampal volume and their FDG PET, say both of them, but no

clinical change, would you be in favor of approving that drug for the treatment of Alzheimer's Disease?

DR. FOGEL: Under the reasonably likely phrase, the answer would be yes you would do that because it would fall strictly under that definition, because it would be a prognostic marker.

CHAIRPERSON KAWAS: Yes. But whatever suggests that altering that prognostic markers makes a difference, is guess what I --

DR. FOGEL: Right. See, the point is that if you saw an alteration that would then take it into the realm of surrogate rather than prognostic marker. And the fact that -- you're saying that this compound actually changes the --

CHAIRPERSON KAWAS: I'm saying, we give you the drug, you've got Alzheimer's Disease, you get the drug, your hippocampus gets bigger now on imaging.

DR. FOGEL: Right. But there's no correlation between the intervention and the outcome relative to the marker. So it still leaves it in the realm of the prognostic marker. And if it leaves it in the realm of the prognostic marker, then under the

reasonably likely phraseology that we're being charged with, the answer is yes, I would vote for that.

CHAIRPERSON KAWAS: Okay. Dr. Van Belle?

DR. VAN BELLE: I would be reluctant to approve because of the two requirements that we need, namely some linkage between the imaging modality and the clinical outcome, and then some information about the imaging modality and the disease progression or disease state. And so at this time I think the imaging work is clearly crucial to studying disease state and disease process. But I don't think we're there yet at the clinical level.

While I have the floor, may I make one small additional comment? Somebody earlier mentioned the situation where the imaging modality would have been significant in a clinical trial and the clinical evidence borderlined. I think it was one of the speakers on the other side of the table mentioned that.

Some kind of analysis of co-variants with the co-variant being the imaging modality might have been one way that the precision might have been

improved and would be based on the assumption that cognition or change in cognition was related to the imaging characteristics. So there are actually ways to deal with this statistically. It's a small point, but nevertheless it might make a trial a little bit more sensitive.

Thank you.

DR. PENN: I think we're seeing here is a shift in a general opinion about where we should take the risks and benefits for Alzheimer's and general neurodegeneration diseases, and that's what the law asks us to do, which is be willing to make a shift towards the risk of putting out drugs that are worse, don't work, don't correlate with the eventual clinical outcomes that we'd like to have them have.

And I think the whole question of what's a reasonable situation in which we would approve such a drug depends upon whether or not we can really find out about that drug in the next X number of years with a Phase IV study that works. Because if we release a drug that has marginal clinical benefit that shows results with a surrogate that we're using, it seems to

me that the only safe way to find out whether that drug really is good is to do what we've been doing, which is the standard type of thing what we've been doing all along, which is require efficacy and safety data over a fairly long period of time. And then we'll have safer drugs. But we're going to miss a number of drugs that we could have approved earlier and found out in a Phase IV whether they worked or not.

So if we have the machinery to check up on what is actually happening in the field after a drug is released, that's fine. But I have my doubts as to whether we have that machinery in hand now to do that. We can require certain things of drug companies and so forth, but try and get somebody not to take that drug or to follow up a double-blind study after it's been released with the impression that it was released because we think it works; it's going to be practically very hard to do. And that, I think, is the real bind that we're having here.

And I think everybody says correctly that we don't a surrogate marker that's been proven to be a

gold standard or that morphs into, as I said, the real thing, which is representing the disease. And that's obvious. But the question is where along the line do we make the reasonable judgment that we can go ahead with safety releasing this drug knowing that this hypothetical drug that works on the "disease process," whether it's worth that risk. And that's a practical question whether the FDA can later enforce the proper studies to be done and be willing to quickly put on an quickly take off something.

DR. KATZ: I'd just say that those mechanism exist technically. In fact, they're requirements if we would approve a drug on the basis of a surrogate that's reasonably likely to predict the clinical benefit, there is a requirement that, as Mary pointed out, that the sponsors perform studies to validate the surrogate in Phase IV.

Now, it may be that from a practical point of view that's very difficult to do in any given case. I think when the regulations were written, it sort of anticipated that those validation studies were well on their way towards being completed at the time of

approval.

Here, at least some people it sounds like believe that those studies need not be underway at the time of approval and whether or not they could actually be done, practically, for the reasons you suggest, I don't know. But there are regulatory mechanisms to require them, technically.

DR. PENN: But if I were a company and I had ten years more to go on my patent, I would take at least 10 years to figure out whether it worked. And I --

DR. KATZ: Well, I think we'd have something to say about that.

DR. PENN: Yes, I know. But, I mean it's not an easy environment in which to deal with. And that becomes -- so it becomes a question of the real specific facts with a real drug as opposed to just sort of generally saying, well we'll accept this or we won't accept this. I think everybody's going to have trouble turning down a drug that makes the hippocampus a lot bigger and clearly in a clean study does, and we have data that the patients certainly aren't getting

worse and it's looking like it's coming into significance. But, clearly, no one wants to just go with a hippocampal drug with no clinical -- what we've classified in the past clinical outcome data.

So it depends on the specific case. And I think until you get a case like that, we can't answer the question in a reasonable fashion.

CHAIRPERSON KAWAS: I guess I want to make a couple more comments.

I mean, I actually think I could not be overly impressed by a drug that makes a hippocampus bigger. Because this disease does not just effect the hippocampus.

I think that in considering this I've been contrasting it with a disease where I actually feel like imaging has a role as a surrogate marker potentially, and that's multiple sclerosis. And there, at least in my concept of the disease, the number and size of lesions can be important in a way that I can more easily see.

Although I know that the hippocampus is part of Alzheimer's Disease, it's not the only part of

Alzheimer's Disease. And so I can very easily imagine that particularly when we're talking about a single marker, that it actually might have the possibility of being actually reasonably unlikely that effecting any single marker will necessarily have a major effect on the disease. Because this is a disorder that effects the entire brain, extracellular, intercellular, almost all the parts, neurochemistry and otherwise.

So, that's the nature of my concerns in approving anything that only effected a single marker.

DR. GRUNDMAN: I'd sort of echo I think that opinion.

I think also in the realm of Alzheimer's Disease clinical trials for demonstrating progression, I don't think that they'd be overly onerous to show some clinical efficacy. I think, you know, we have instruments. We have the CDR sum of boxes. We have the CGIC. We have the ADAS Cog. We know what their rates of progression are over, say, one year or 18 months. And I think we can design trials to demonstrate a one-third decline or some clinically relevant efficacy outcome measure and see whether or

not the imaging supports that clinical conclusion without using the imaging marker as the sole criterion.

Now, it gets more complicated when you move to prevention. Because there, I think, it requires much larger sample sizes and smaller effects.

I'll stop there for the time being.

DR. WOLINSKY: I think it depends -- I've got to be careful with words. So we're not talking about surrogates anymore. We're just talking about anatomical measures of disease or biochemical measures of disease?

DR. KATZ: Well, I'd call it --

DR. WOLINSKY: Because we had to throw the surrogates off the table?

DR. KATZ: No, no. I'd call them unvalidated surrogates.

DR. WOLINSKY: Okay. Okay. Became semantics could really get us into trouble here if we're not careful.

I would say I actually wouldn't -- if I was in your seat and somebody came to me and said

okay, I want to use one of these markers as a primary outcome measure and I want you to be able to tell me in advance if I won on that, that you'll give me approval even if I lose on my clinicals, I would say fine. Pick atrophy because I don't think we're going to get a brain Viagra that's going to blow up the size of the brain in about 15 minutes. So that you would actually have a shorter term study that would not allow you to actually get a clinical correlate.

Now, it might be a little bit more uncomfortable if you pick something like PET scanning, depending on what the law again was, or if you picked spectroscopic marker. Because there may be other things that could begun to normalize or reverse those kinds of changes. And in a short term study you might see a change in that that might or might not have a clinical correlate.

So I think, again, it becomes the practicality of what that marker is, what the expected time course is, how long it's going to show and how likely is it to be linked with the clinical outcome at least to a near miss level.

So I don't think that this is as scary as it sounds, just because of the nature of the technology right now.

DR. CHIU: I still strongly believe you measure modality, you know -- better than MRI or PET scan combined. You cannot use CT, you cannot use x-ray. And we don't treat patient AD just based on the MRI findings. We need clinical, you know.

CHAIRPERSON KAWAS: Actually, I think that that's Dr. Katz' question; Should we consider approval a drug that effects the imaging of whatever sort or however many different image modalities you choose, but does not have a clinical outcome.

DR. CHIU: Okay. Single MRI imaging, you cannot judge from that. You need to see with examination. You cannot just judge one. If you do blind study and you treat a patient as AD, but you have it today, you have a --- 6 months, you're able to see the effect of the medication. You can just from one study.

So to me I think what else can you do? You don't have any other modality, imaging modality.

So I do believe this the most powerful imaging modality.

CHAIRPERSON KAWAS: So you would support a drug approval based on only longitudinal imaging studies without clinical demonstration of any clinical changes?

DR. CHIU: Definitely. What I need to see is a study, not just single.

CHAIRPERSON KAWAS: Okay. Mr. Ramsey?

DR. RAMSEY: I'll give the short answer. I would not approve a drug. That would be too much of a leap of faith for me to say that it would be likely to have an effect when, in fact, you can't see any effect. And I would defer to Jack Welsh, who I know is a little bit out of favor. But one of his sayings was "to accept reality as it is, not as you wish it was." And that's what I would say.

So, no.

CHAIRPERSON KAWAS: Thank you, Dr. Ramsey. Your brevity especially is appreciated.

Dr. Beam?

DR. BEAM: I guess my answer is a question

for Dr. Katz.

Dr. Katz, if this were to happen, what would go on the label as that drug as far as a claim for what the drug does?

DR. KATZ: Well, I have no personal experience with approving drugs on the basis of surrogates. It might say something like decreases hippocampal atrophy, or something along those lines. But we wouldn't approve it in the first place if we didn't think that that meant something clinically. In other words, reasonably likely at the least to predict a clinical outcome.

So what exactly the language would be, I don't know. But we wouldn't approve it, as I say, in the first place unless we thought that language was reasonably interpreted to mean it had an important clinical effect.

DR. BEAM: Well, if the language said simply that this maintains volume, for example, and doctors believe that this is associated with positive outcome for Alzheimer's patients, something like that, I think I could go along with that approval process.

DR. KATZ: Again, the specific language is -- it's certainly important, obviously. But the fundamental question is whether or not we ought to approve it in the first place.

Drugs can do lots of things and effect lots of outcomes that we can measure, but we would not necessarily approve a drug on the basis of its effect on some serum marker or something if it didn't mean anything clinically. So we still have to bite the bullet, in effect.

DR. BEAM: Right.

DR. KATZ: I mean, we have to decide first and foremost fundamentally whether or not the effect is, again, reasonably likely to predict something important clinically. I mean, that's -- we can't get out of the conundrum by just describing exactly what the drug did, let me just say that. We have to believe that that meant something.

DR. BEAM: Then my answer would be no to that question.

CHAIRPERSON KAWAS: Thank you, Dr. Beam.

Dr. Wolf.

DR. WOLF: I sort of had addressed this question before. The answer is that if there is a positive indication from the imaging and while there is no conclusive clinical indication of progression, there's no regression, there's no deterioration, then the drug should be considered.

DR. SORENSEN: Okay. So I'd like to give my answer by telling a little story that I think I found in literature that seems to support a scenario under which I could answer this question.

And that's the drug called Etanercept. As I understand it, it's a drug that acts on rheumatoid arthritis. And it was originally compared to a standard treatment for rheumatoid arthritis called methotrexate. And it had both the clinical outcome and it had some imaging outcomes.

And the clinical outcome was to measure some kind of rheumatology score every month and the imaging outcome was to image the joint space narrowing and erosions at 6 and 12 months.

It took off and looked really good in the first few months, but its primary endpoint at 12

months it just barely missed statistical significance. The joint space narrowing didn't work, but the erosion scores showed a really dramatic improvement in erosion score reduction. So it looked like the knee joint was looking a lot better. And it was approved. I don't know exactly how, I wasn't privy, but it did get approved and they had to do some follow-up studies. And the follow-up study was published this year, and that was some 2 year data which was basically the same protocol. And in 2 years the drug did work. It showed that the rheumatology score, the clinical outcome score was better and the erosions continued to be improved and overall the imaging endpoint just lead the clinical endpoint.

And this was a scenario where I think people understood the pathology. Not all of the imaging endpoints worked. The joint space narrowing was clearly not working. So that biomarker failed. But one that people did understand and seemed to fit was reasonably, and I think in hindsight I would say, and I think most people would say, that the agency made the right decision there. They got a drug that

was effective onto the market sooner.

And I would say that if a similar situation came up with Alzheimer's, that I would be in favor of the same course of action. If you just barely missed your clinical endpoint but there was some earlier data that suggested that it had worked, and the imaging was compelling. Maybe not all of the imaging endpoints, but at least some that made a lot of sense. Then I would say, yes, the biological link between -- maybe not just one endpoint like, you know, the hippocampus, but if the whole brain or if it were the ventricles, or if it were glucose. You asked us which one of those, I know, and I'm waffling a little on that. But I would probably say whole brain volume for me and probably glucose metabolism.

If those were to improve, I'd say that that would be a compelling story and I would seriously consider it.

CHAIRPERSON KAWAS: So is -- let me understand your answer. Is your answer that if the imaging is positive and the clinical is borderline, you'd go with the imaging.

DR. SORENSEN: That's correct.

CHAIRPERSON KAWAS: If the imaging is positive and the clinical was negative, what would you say?

DR. SORENSEN: If the imaging was positive and the clinical was clearly negative, say, that there were no indication that -- or that the patients -- well, let me break it down.

If the patients did no worse than the placebo, I would probably still go with the imaging. IF the patients did no -- did worse than the placebo, I would certainly not go with the imaging. I don't know if that answers your question.

It'S a little a bit like one of the earlier respondents said. It's tough to argue this in the abstract. You'd like to actually see a case in front of you.

In the case of Etanercept, it was a tough call because the primary endpoint wasn't made, but it was borderline, and that made it easier.

How borderline is borderline before you'd call it? Well, I'd have to see the case or I have to

see the situation.

If it were far from borderline, I probably wouldn't go with the imaging.

CHAIRPERSON KAWAS: So maybe the best way to phrase your answer is that you think imaging should be an ancillary, but do you think it should be primary?

DR. SORENSEN: I mean, it would be more than ancillary. It was more -- I mean this drug got approved on the basis of the imaging and not on the bases of the clinical. The clinical didn't work.

CHAIRPERSON KAWAS: So does that mean imaging should be used as a primary outcome then in your opinion?

DR. SORENSEN: So in my opinion I think they made the right decision here by using imaging as a primary outcome. They did. I don't know whether it was accelerated approval or not. Like I said, I wasn't closely involved, but I just read the literature. But I think they did get approved and they did have to do some follow-up studies, so maybe that means they were given accelerated approval with

these conditions.

And that kind of scenario is one that I would endorse for Alzheimer's as well.

CHAIRPERSON KAWAS: Okay. Dr. Kim?

DR. KIM: I think my hangup is more of making the connections, more definite answers. I see evidences and yet that evidence doesn't really show me whether that has a direct correlation with the AD at this point.

With that said, I think what I'd really like to see with the present technology would be more of a Phase IV studies which also give us a little more for us to understand, for the technology to catch up even better. And have a lot more data.

I think one of the problems that I have is that we don't have enough data in whether it is one modality or multiple modality, or even the normal ways that we haven't even thought of yet.

So I think it's more of which came first or which comes first. But I would like to see this in one of the Phase IV and come and revisit this portion here. So even with the present data, I don't agree

that I would approve based on this.

CHAIRPERSON KAWAS: Okay. Thank you very much.

Mr. Perez is keeping a tally of the votes of some sort, and I'm glad it's him and not me. Because I'm not sure how some people voted.

But, Dr. Katz, did we help in any way with these answers?

DR. KATZ: I, too, am having a similar difficulty.

The no's were pretty clear. I'd have to add up how many I had. But maybe it would be useful to further probe the people who felt that a drug currently could be approved.

There are a few questions that I have that would help clarify people's thinking for me. I mean, I think I understand when people said no I don't think we're ready; I understand that. It would be useful to have the deeper understanding of exactly why the people who said they might or they would, thought that way. It would just be helpful to us to understand that.

So there are a couple of things I'd like to ask. I don't know if you want to do that now or -- okay.

The first is, the one that I did raise, which is the question of which marker. The folks who think we're in a position now to approve a drug on the basis of an effect solely on a surrogate that's reasonably likely to predict clinical outcome, which marker or which modality or modalities, which imaging modalities.

The other is, there are two other considerations I think we ought to talk about. Again, the purpose of using a surrogate in this case would be to do shorter studies, smaller but also shorter studies. So I have two questions.

One is what about the possibility that an effect that you might see, let's say at 3 months on whatever surrogate you choose, what about the possibility that that might be transient? In other words, the understanding here is that if an effect was seen, it would presumably persist and ultimately translate into a clinically meaningful difference to

the patient. But suppose that in 3 months you see an effect on the surrogate, but at 5, 6 months it's not there anymore. I mean, we wouldn't have that information at the time of approval on the basis of, let's say, short term studies. So I wonder what people think about that.

The other reason that people like surrogates is because they can be very sensitive, so you need fewer patients. But what about the possibility that a statistically significant difference on whichever marker you choose is seen in a relatively short term study; what about the possibility that even if that effect persists it might not get any larger and how do we know that that size of a change actually ultimately will translate into a clinically meaningful outcome?

So, which marker or markers, which surrogate or surrogates, how do we account for the possibility that the effect that you see in a short term study may just be transient, and do you worry about that? And the possibility that the very sensitive measure will pick up very small changes,

let's say, in hippocampal volume and the possibility that that in fact might have no clinical meaning?

So, it's a lot, but it would be very useful for us to hear what people think about that.

CHAIRPERSON KAWAS: Okay.

DR. KATZ: Those of you who said no, relax.

CHAIRPERSON KAWAS: Although we have opinions on those things, too.

Now I'm going to let the people who said yes start self-identifying. My take on it is it's more of this side of the room, but if I'm ignoring anyone over here, I'll make sure.

Oh, all of a sudden -- I got to tell you, Dr. Katz, all of a sudden the vote to me looks like it shifted. A lot of people want to talk now. But let's start over this way.

DR. WOLF: Okay. Let me start here.

First of all, any imaging modalities we have, we would be using, we would have by now long term studies available so that we know that we would have some feeling whether something seen at 3 months

maintains itself at 6 months, 9 months, etcetera. These are part of the studies that are ongoing, so we would have some background on what those changes represent.

CHAIRPERSON KAWAS: Well, maybe the question that he would like you to answer then is how long do we need to follow before we should feel confident that it's not transient?

DR. WOLF: I am happy I don't have to make that decision. But let me give you an example of the kind of area I work in, which is oncology. And in the case of tumors, the gold standard is reduction in tumor, in solid tumors reduction in tumor volume.

Now, you can have a tumor that is stable which is composed of dead tissue. And in those cases functional measurements of metabolism and perfusion give you much more valid information and much earlier information, the changes in volume of the tumors.

So that it depends very much what is the biological parameter we are measuring and what is a sequential basis of those biological changes appearing.

So I don't know enough about Alzheimer's Disease to know how it progresses, what determines what, what are the different biological steps that will cause neurodegeneration, etcetera. But the question is if we have a marker, an unvalidated surrogate marker that gives us some information that suggests that a positive change is occurring and if there is no clinical indication that is contrary to that, then we have a reasonable probability that something positive may be happening and it is worthwhile considering.

CHAIRPERSON KAWAS: Dr. Wolf, perhaps you could speak directly to Dr. Katz' question; which marker do you favor, if any, right now is the first one. I mean, is there any particular one that grabbed your attention or just the idea of markers in general in neuroimaging?

DR. WOLF: I would say at this moment the idea of markers, because I am not -- I have listened to some of the information we have, all of them give a limited information of what's happening at the tissue level, but they don't give a good comprehensive idea.

So I think it would be a combination of different markers that would have to be considered.

CHAIRPERSON KAWAS: Okay. And you hedged your bet sort of on the issue about transient effect that may washout over time. But how about his final question; the possibility that these images may be so sensitive that they're detecting small changes that may not be clinically relevant?

DR. WOLF: If they detect small changes, then they're likely to be washed out by comparing a number of different patients. Because you would have-- I mean, you would not have the same -- if you had exactly the same degree of every comparable level of small changes, then that would be meaningful. Just statistically I would suspect small changes would disappear in the analysis if they're compared over a sufficient number of patients.

CHAIRPERSON KAWAS: Okay. Thank you.

Who wants to -- Dr. Sorensen?

DR. SORENSEN: I think -- I'm sorry.

CHAIRPERSON KAWAS: No, I was just going to sort of remind the questions; which marker, the

possibility that the effect might be transient and the size of the effect.

DR. SORENSEN: Right.

I think that none of the studies -- or none of the markers that I've seen have been -- have followed the rigor that the FDA or any good scientist would require them to follow in collecting and analyzing them.

I mean, I'm familiar with Dr. Love's draft guide document on medical imaging, and as far as I can tell, maybe there's one MRI study but none of the PET studies and none of the MRS studies have followed that level of rigor in using centralized readers, in standardizing their protocols and all of that.

So, as a result, I'm not sure that any of these in their current state today are actually acceptable as an endpoint. And my earlier positive, you know, views were assuming that somebody could come and pull that level of rigor together and actually follow the rules, you know, the scientific rules that exist for doing science well.

And so as a result, that's influential

because the things that are closest to that I think are the MRI volume measures. They seem to be the ones that have been used a couple, or at least one multi-center trial with a prospectively defined analysis algorithm and endpoint.

And so I would say that the volumetric markers are probably at the top of my list just because they seem to have the best track record for being analyzed.

The sensitivity thing, I think most people like it's not just the sensitivity, it's the lower variance that makes these interesting. But I think the point's the same, that you might -- if it's a low variance whether it's sensitivity, you see things that might not be clinical relevant. And I think that is a very relevant issue if you saw a -- you know, an effect size that was too small or that seemed, you know, within one standard deviation of the noise of your measurement, that wouldn't be very interesting to me. And I think each of them have different denominators, so I don't know the right language to use to describe how much of an effect I'd expect to

see. But even with the lower variance if I saw two standard deviations of an effect, I think that would make me feel more comfortable even if that was small relative to what I would expect the clinical outcome.

The transient thing I think is also a very important point. And I think that that has to be measured against the background of the natural history of the disease and it's variation. And from looking at the graphs we saw today, I would guess that at least a year you would have to see this effect. That if it were less than that, it would be suspect.

CHAIRPERSON KAWAS: Thank you.

Dr. Chiu, do you want to -- I think you actually have already told us which markers. I think you told us MRI and --

DR. CHIU: Depend on from the common more point of view, you know, those are static image you're able to see the hippocampus, the size of the ventricle. Now we have a new pulse sequence called a T1 flare image. You're able to clearly see the structures of the -- and so forth.

So the MIs getting, you know, more and

more to more powerful --

CHAIRPERSON KAWAS: So any particular MRI measurement that you think or multiple measurements of anatomical MRI?

DR. CHIU: I think it's all individual. We have a computerized, you can use it. I'm talking about really talk about the imaging of the grading signal and noise, it come with T1 flare. It really give you very crispy gray and white matter. You really can see it.

Earlier mentioned about the T2. We can do T2, just make sure we're not dealing with a multiple-- dementia, you know, that kind of -- it's helpful.

MRS spectroscopy, not many of these center you have that MRS. But that's another powerful tool. It take longer time. Usually you have to take good time to measure that.

Come to the PET scans, I don't know if its approved or not. We do a PET scan, but that's more expensive.

So I'm talking about standard imaging, T1 weighted coronal view plus volumetrics that's can be

done. I think that's probably the more --

CHAIRPERSON KAWAS: And do you have any concern that some of these effects may be transient and/or too small to be clinically meaningful?

DR. CHIU: Yes. Back to the MS, I don't know, to way back to ten years ago we have a MS study. We use a standard T1 and -- T2, we're able to make diagnoses. But as you said, it's come and go, not that specific.

Then we have this MP -- and you can pick up more earlier, more definite MS. Now we're using this as a gold standard. A lot of patient come in clinical and questionable MS. And we're able to make a diagnoses window there. Do all kind of spinal tap I'm able to make a diagnoses. We are the one who give them more definite diagnoses.

So come to transient, I think clinical doubt you probably have to do every 3 months to see that how that changes. They might -- patient maybe dehydrated or alcohol taken -- show the changes. But at least over a year and every 3 months for -- the symptoms, and that's how we do now.

CHAIRPERSON KAWAS: Thank you.

I think there was some interest on talking at this side of the table, too. Dr. Fogel?

DR. FOGEL: Yes. Well, I agree with Dr. Sorensen that none of the data that we've been shown today reaches the rigor that we need. And when I had mentioned the MRI volumetric analysis and the FDG, I meant it in combination with prospective future trials that would be multi-center and wide scale trials targeted things like MRI volumetric information as well as functional information like FDG.

In terms of whether or not the effective short term and transient, I agree about the year definition. I mean, in the accelerated language it says that the surrogate has to be reasonably likely to predict clinical benefit. And I don't think anything that's going to be transient or short is going to be reasonably likely to translate into clinical benefit unless it's present for a long period of time.

So, a year is better than 3 months, so I would pick a year although I don't have data on that.

CHAIRPERSON KAWAS: And what would you say

about the data that we saw today, the MR spectroscopy data in the randomized trial that showed an effect on the imaging that washed out before the study was done, which was 24 weeks, as I recall?

DR. FOGEL: That was transient, I would think, because it did wash out. And since we don't have it --

CHAIRPERSON KAWAS: And that, by the way, was in a study where the drug did show clinical benefit, by the way.

DR. FOGEL: That's right. That's right.

CHAIRPERSON KAWAS: So in this case the clinical benefit continued to the 24 week and beyond mark, apparently, but the MRI spectroscopy which looked like it was correlating with clinical benefit disappeared by 24 weeks. I mean, how would you interpret that sort of data?

DR. FOGEL: Not really clear how you would interpret it with the exception that it might have done the -- I'm not really sure how I would interpret it.

DR. SORENSEN: The graphic showed how the

ADAS Cog score is the same at week 24 for placebo and Donepezil. So I think they had equilibrated by the end.

DR. FOGEL: So you're saying that it didn't show clinical benefit.

DR. SORENSEN: It did early on.

DR. FOGEL: No, no, no. But I'm saying it was transient because it didn't show it for a sustained period of time?

DR. SORENSEN: It did for 24 weeks, but not for 30, at least the graph I'm seeing.

DR. FOGEL: Right. And if we use the definition of --

CHAIRPERSON KAWAS: But the MRI did not show it at 24 weeks.

DR. KATZ: But that was after washout.

CHAIRPERSON KAWAS: Thirty is washout, so that doesn't count.

DR. KATZ: Right. But it does suggest that -- and I think we ought to talk about this later for the people who said no, so you can't relax completely. I still have some questions, but the fact that that

surrogate after washout, after discontinuation of the drug, was back to where the placebo patients were suggests-- again, there's a suggestion that if you -- one way, one operational way to show that a drug has an effect on progression independent of imaging is to take the drug away. And if the effect still persists, that's pretty good evidence that you had an effect on progression.

In this case, the effect on the surrogate didn't persist. The drug was discontinued.

CHAIRPERSON KAWAS: It didn't even persist while the person was still on drug in this case.

DR. KATZ: Well, okay. Even worse, right. But even if it had and you took away and it goes back after a drug is discontinued suggests that that's also documenting just a temporary symptomatic effect and not a structural effect.

DR. FOGEL: Right. So it seems that there's some debate about that. But, I mean, I would think that you would need a year to actually show a sustained clinical -- a benefit that wasn't shown before but that might have actually had a clinical

benefit.

And in terms of the imaging modality being too sensitive to small changes to make a clinical difference, I think that that's a very important point. And the reasonably likely phraseology, again, says that it needs to predict clinical benefit. And if one feels that the changes are too small to predict clinical benefit, then it shouldn't go into the approval process.

But the other thing I want to just bring your attention to the fact that there are -- we might have other surrogates that we hold to a higher standard that you need larger -- that aren't as sensitive that you need larger changes and we risk missing efficacy because we need to see those large changes. Because those surrogates aren't sensitive. And so we basically balance in our equation everyday in other surrogates that are less sensitive than imaging studies -- danger of missing efficacy.

CHAIRPERSON KAWAS: Thank you.

Dr. Penn wants the last word before the break, so take it away.

DR. PENN: At least a year, probably 2 years overwhelming evidence that the surrogate moves, not equivocal evidence. And no deterioration clinical state during that period of time. So the criteria for a fast release drug without the usual clinical benefit shown yet has to be enough so everybody nods, yes, that really is the right thing to do. So it has to be very strong evidence.

DR. GRUNDMAN: Just a question. You know, if you're going to do a study for 2 years on Alzheimer's Disease and require that the surrogate effect persists for that length of time, wouldn't you expect to see a clinical outcome measure showing the same effect over that period of time if you powered the study adequately? It's really -- I think it's a practical question and if over that length of time you didn't actually see a clinical benefit, even though you saw something on the MRI, I'm not sure how you would interpret that.

CHAIRPERSON KAWAS: Okay. I mean, I think what the last two speakers have both said was essentially that we should do the studies as we

normally do them and look at the surrogate markers in addition to the clinical markers? Isn't that more or less what I heard?

I mean, I think a lot of interest in doing these surrogate markers is the idea to not have to do two year studies. It's the idea that if I can give something to someone and make their hippocampus go up 20 percent in size, there's something that maybe that can happen quickly and that can give me information quickly that I can then use to shorten the studies, not only in terms of numbers but certainly in terms of time. But at least some of the people who I couldn't tell how they voted the first round, really are saying that they want to see them together before they would feel comfortable.

Is that true?

DR. PENN: That's what I said, I'd be satisfied in a year or two. What I said was that I'd be satisfied by overwhelming evidence maybe in a year without definite clinical changes for releasing it with a Phase IV coming on after that.

CHAIRPERSON KAWAS: Well, I guess what I

meant, though, is you would want the clinical markers also being measured at the time? You don't want just the --

DR. PENN: Oh, yes. I don't think anybody's saying we just have to go to surrogates and forget about clinical results.

CHAIRPERSON KAWAS: Can we quote him on that?

DR. KATZ: Well, no, I think there's a difference between measuring the clinical outcomes and requiring that they be positive by whatever definition.

DR. LOVE: What I am hearing is several different approaches which tend to be leaning towards what you're saying. Look at the clinical, make sure that it's not deteriorating or at least that there's some level of static, and then look at the image.

What I'd like to know, one thing I've heard and whether you want to take the break and come back and address this, it has more to do with question one, is related the how would you validate or at least the how you would be comfortable that it is meeting a

reasonable standard for surrogacy.

And that is, to go back to your discussion earlier, you mentioned composites. Several people said anatomic plus some type of functional measure. I would like to hear some discussion on how you would look at that. Would you be at this point ready to make some type of assessment on what you want to see for the anatomic, what you'd want to see for the functional; does one have to lead the other in relationship to the clinical or not. Is there enough information to move to that yet. Are you just looking at coprimary, meaning you'd have to see a change in both measures or one of the other. Just some discussion on that maybe after the break.

CHAIRPERSON KAWAS: And with that, I think we will take a 15 minute break. We'll be back shortly after 4:00.

(Whereupon, at 3:47 p.m. a recess until 4:07 p.m.)

CHAIRPERSON KAWAS: Okay. We need to pick this meeting with Dr. Love's questions. And just to sort of refresh where we were, I think overall the

Committee has a lot of enthusiasm and positive feelings about the potential for neuroimaging to be used as a marker in studies of Alzheimer's Disease. And Dr. Love would like to know if we were to go about getting the data or the kinds of studies that need to be done in order to allow us to use neuroimaging effectively. And some of her specific questions included do we need an anatomic measurement only. Do we need a functional measurement also. Do both of them have to be positive. Does one of them have to lead the improvement or not.

And so the table is now open for the questions that Dr. Love posed to us. And who would like to start with this challenge?

Dr. Van Belle, a statistician.

DR. VAN BELLE: Oh, I just want to say that I have no opinion on which modality should be used. But I think what I would like to say up front is that we would agree, I think, that the context has to be that of randomized controlled clinical trials. That observational studies won't do it. I just want to make sure that we understand the game plan before

we go talking about modality.

Thank you.

DR. LOVE: Just maybe to clarify my question. My thought was that in the context of that randomized trial what I'm hearing is that there are persons around the table who are interested in some one or more and the idea or a theme of a composite has come up several times. So I'm curious how you would go about working that into the study, what specifics would you be thinking about.

CHAIRPERSON KAWAS: Does Dr. Fogel want to suggest a composite?

DR. FOGEL: No.

CHAIRPERSON KAWAS: I mean, I'll start by saying I think that most of the discussion seems to me to coalesce around the idea that two instruments would probably be better than one, and that most of the people who have suggested combining have generally suggested one anatomic and one functional measure.

And it seems to me that overall on the table the most common anatomic measurement has either been hippocampal volume followed by total brain

volume. And the most common functional measure, I believe, that seems to be surfacing is PET scanning.

Now, is there anybody who agrees with any of those or disagrees and wants to -- Dr. Fogel?

DR. FOGEL: No, I agree that one would need -- I think that you would need both, and that I think that Dr. Love had questioned whether or not we needed one positive or both positive. And I would hold that you would need both positive and, as you mentioned, that there should be one anatomic like hippocampal volume and one functional like PET scanning. And they'd compliment each other because as people had rightly suggested, if you give a drug and it causes inflammation and edema and doesn't show a shrinkage, that you would anticipate wouldn't show any functional benefit yet the volume would be there. So, I mean, you'd need both, if you will, positives to give you a more reasonable likeliness, if you will, to show a clinical benefit if the prognostic marker is not transient and stays there for a long period of time.

DR. WOLINSKY: If I might, I guess I need

some help from the experts over here that might be doing those PET and anatomic studies together. Because I would like to know before I built a composite that the measures of the composite were at least somewhat independent of each other so that I wouldn't be just measuring the same thing twice. And I'm not exactly sure that I heard from the experts who were presenting what the interdependence of those measures are where the groups have tried to look for that.

So that would be very important, especially since in my -- as best as I can remember, back to neurobiology 101, the number of cells there determines the amount of metabolic need determines the amount of cerebral blood flow determines the FDG. And all of that could just be a function of atrophy.

DR. JAGUST: Well, you know, so the empiric answer to that is that one can correct PET images for atrophy using the MR data. And you can do it in a actually quite sophisticated way taking into account the thickness of the cortical gray matter and so forth. And when you do that you still find substantial differences between Alzheimer's patients

and controls, although those difference are attenuated.

Now, I think you can always push that question further and further down stream to the fundamental disease mechanisms and ask are the same fundamental processes causing the changes in metabolism and the changes in brain structure. And, of course, you can only conjuncture about that.

DR. WOLINSKY: The problem is I'm not asking whether or not you still see fundamental differences between controls and patients with Alzheimer's Disease. I'm asking whether we're in a longitudinal study those move in an identical fashion or whether they move in a differential fashion. And that question is critically important if we're talking about using one or two measures in a therapeutic trial as either a supportive for a surrogate marker.

DR. De CARLI: Yes. I guess that Cliff's not here.

The one thing I think that has been done where we see differentiation of those two effects is in vulnerable populations where to date MRI data,

structural imaging has not shown changes where metabolic imaging has shown changes. The problem is that like the E4 carrier, is that -- I'm going to turn it over to my colleague -- about what the outcome of those individuals has been. I mean, do they progress on to dementia without structural imaging or is this -- and so how it relates to as an early marker. But other that, that's the only evidence that I know about where the two are disconnected. Most the other evidence suggests that they're connected. But almost all of it's cross-sectional. We have some longitudinal data, but we haven't analyzed it completely yet.

DR. REIMAN: I would concur with the idea that the PET changes one sees cross-sectionally are not entirely attributable to atrophy and that the data is available but hasn't been looked at to determine the extent to which atrophy accounts for those changes.

And I would also concur with the idea that we can see these PET changes at the moment in distinguishing people at risk for Alzheimer's Disease

from those who aren't prior to the onset of memory impairment, with a little bit more power with imaging with PET than with MRI.

I think the rationale for using the complimentary measures is that it is very unlikely that a confound, say, on brain swelling will effect a change in neuronal activity and vice versa. And the advantage of using both is that you can reduce the chance that you're going to see that kind of confound.

CHAIRPERSON KAWAS: Thank you.

I think that, you know, what has been addressed the table so far is the idea the two studies is better than one; you want them both positive and you want them both independent.

But I'd like to suggest the possibility that I can easily imagine a legitimate marker and a therapy influencing only one of those markers and not both of them. And I wonder if we're not mixing up our apples and oranges here if we pick out two or more markers and insist it be positive on all of them.

I mean, you can change with drug the metabolism of the brain potentially and change your

functional outcome measure, i.e, your PET or whatever, and potentially not change your anatomic and still have something that's very efficacious.

So do we really want to insist the two measures which are absolutely necessary and that one should be functional and one should be anatomic?

DR. WOLINSKY: Well, I thought the question was a question of a composite. And so I was constructing a composite and if they were all doing the same thing, then I have no reason to waste my time, effort and money on a composite when I can just measure one.

CHAIRPERSON KAWAS: If you knew which one worked?

DR. FOGEL: Well, just like any other test, the composite has a sensitivity and specificity. And because it's not perfect, you're going to miss some efficacious drugs and you're going to let in drugs that aren't efficacious. And so just like any other test that is sensitivity and specificity, but the bottom line is that the accelerated phraseology wants us to have a reasonably likelihood to predict

clinical outcome and so we want to error on the side of being able to let a drug out that we're sure that to the best of our ability it can predict and to a high likelihood that it might predict a clinical benefit. And to do that, it would seem logical that one would want to have both an anatomic parameter as well as a functional parameter to do that. And we're trying not to let drugs in to the general population that may not prove clinical benefit by doing so. So we're erring on the side of leaning towards the clinical benefit at the cost of not letting a drug out into the general population that won't have a clinical benefit and may cause adverse effects that we really shouldn't have let out in the first. So we're really trying to increase that likelihood as much as possible, which is why you would want two at the exact same time, two simultaneously having a positive result in this composite, which is why it was suggested in the first place.

DR. SORENSEN: Dr. Kawas, I don't think you need both. I like Dr. Penn's point. One compelling story is good enough. And I agree with you that we, as

much data as we've seen today, I think everyone would agree we're still at a fairly early stage of our understanding of these markers. And I wouldn't want to insist that a drug had to succeed at both this early in the game.

I think the biggest challenge around all of this is that we still don't have really enough data to speculate about specific details. So to be explicit with Dr. Love's questions, my own sense is that given the numerous single center studies of both PET and MRI, I feel like when someone gets around to doing the multi-center trial the right way, that there will be a link between the pathology and the imaging. But those links have not been established today, to my knowledge, in a well designed, well controlled prospective trial. That data just isn't there.

And so it's hard to say which one would be first or which one is better. I think we'd be speculating and you're hearing some speculation at this point, but it's speculation.

If the rumors about the NIA sponsored study or these others, maybe they're included in

commercially sponsored trials are true, maybe within a year or so maybe we'll have enough data, hopefully well before somebody would come to you or at least concomitant with when somebody came to you, and then we could use that to help guide this.

I'd be nervous offering guidance to somebody right now when a well designed study could change that guidance.

DR. PROVENZALE: Comment. I'm in agreement with Greg. I think we're missing a lot of the basic data. It's similar to feeling an elephant from many different angles.

The data that we've seen and that we've read in the literature is very promising. When I mentioned correlation of, let's say, PET and volumetric MR imaging or PET and spectroscopy, I was basically pointing out that we don't know what the correlates, we don't know what the glucose metabolic rate change is in areas -- I mean, I don't think we do. In areas of hippocampal shrinkage accounting for volume loss or in, you know, what's the correlate of NAA decline with glucose metabolic changes on a pixel

by pixel basis? We don't have that information.

I think a lot of what we're basically talking about here is we're not really answering the questions of how would you design this from the FDA perspective. We're really kind of outlining a wish list of the necessary, from our standpoint, perquisites for moving forward. And there's time to do this before, like Greg said, a drug comes to market. But these are the things that we'd be interested in learning more about before we could answer your question.

CHAIRPERSON KAWAS: Dr. Wolf?

DR. WOLF: I think we have three levels we have to worry about; biochemical changes, physiological changes and anatomical changes. And I think one of the areas, again adding to the wish list you just mentioned, that we need a lot more information is markers of molecular changes that occur relatively early and that can be measured and that can be indicative of what's likely to happen later on in the physiological and in the anatomical phase.

So the question really -- the answer to

your question is we don't know at this moment which one of these measurements is the most efficacious one, but I think we need to accumulate the data and see which one correlates best, and hopefully try to develop some additional markers that are more specific and go more to the mechanism of the disease process in order to really have a good handle.

DR. OLIVA: I would like to hear from the Committee members who earlier said that, no, that there are no markers that are reasonably likely to predict clinical effect. I'd like to hear your answer to the following question: Dr. Fox earlier this morning suggested to me, anyway, that quantitative MRI imaging might be reasonably likely to predict a positive clinical outcome if that effect persisted. So what would you think of a clinical trial design that would incorporate quantitative MRI imaging as a primary outcome that also incorporated a randomized withdrawal design that showed persistent effect?

CHAIRPERSON KAWAS: I'll start, because I think I'm one of the people who suggested -- well, your words were no markers likely to predict outcome,

and actually that is not what I think.

I, in fact, think that it is very likely that some of these markers would be relevant for outcome. The problem is I've not seen the data to make me know that. So I think there's a bit of a difference between those two things.

The second part of your question was then what would I think about a design that withdrew. And here I think it really depends on the mechanism that we're trying to get at with a particular drug.

So although I understand why everyone's suggesting a composite, I think that's too stringent a test. I think a drug could easily work in a way that would show up on a functional surrogate marker and not show up on an anatomical surrogate marker and still be absolutely relevant to the outcome. So I actually don't particularly like the idea of a composite anything, because all you're doing is putting together a bunch of things in the hopes that somehow that makes you closer to right.

But the part that's missing with the single is the information that would make me feel

confident that if I make that change on the PET scan, that I've also made a change in the patient down the line. And in that sense, the design wouldn't help me at all I don't think.

Dr. Fogel?

DR. FOGEL: No, a composite doesn't actually do that in the hopes that you're going to be right. What a composite does is it takes the probability that you're right on two of them and gives you a higher probability that together if when they intersect that you will be more reasonably likely to eventually predict an outcome. So it's not that one is hopeful. You're trying to be more specific because when you're saying -- if you don't have data to show that the unvalidated surrogate is going to have a clinically relevant outcome, then you have to hold it to a higher standard. And to hold it to a higher standard, you have to be more specific. And to be more specific, you need to have more than -- it seems to me from the data that has been presented, that you need to have more than just one unvalidated surrogate to be positive simply because it will be more specific in

terms of effecting the disease and more likely to actually have a relevant clinical outcome.

DR. KATZ: Yes. I'd like to just expand on Armando's question. But I think it's the next critical series of question or series of questions that I think the no voters should address, which is what if anything at the moment should be part of the design of a study that would allow us to conclude that a drug has an effect on the underlying progression of the disease?

Dr. Fox had talked about including a withdrawal phase, as Armando pointed out. You know, obviously, people have talked about the so-called randomized withdrawal or randomized start clinical trial without surrogates, which involves at some point withdrawing the drug or putting people who had not been on the drug on the drug and seeing whether or not any effect persists between the drug and placebo patients. So that intrinsically has a randomized withdrawal phase. And we believe that study is interpretable as having an effect on progression if the effect at the end of the randomized period

persists during the withdrawal phase.

So I'd like to hear what again, if anything, elements ought to be included in the clinical trial that will allow us now to conclude, if anything. If you think this is even doable now, that if a drug doesn't effect on progression.

I mean, Michael talked about a study in which simply just correlated the clinical with the surrogate at some point down the road, and that might be sufficient.

So, I think we really need to hear from those of you who voted no. We're not ready to rely solely on a surrogate Are we ready to rely on any combination of elements in a trial and what ought those elements to be to support, in effect, a claim for an effect on progression?

DR. VAN BELLE: Well, as the most no of the no's, I suppose.

First of all, let me say that I would set very high hurdles for surrogates. I think there's enough evidence and there are enough bad cases that I think the agency should move very carefully with

respect to accepting surrogates.

Secondly, if I were in charge, what I would do is I would still take a randomized clinical trial with the endpoint as the primary outcome and a surrogate as the secondary outcome. And whether I did this in terms of a randomized withdrawal or one or the other designs, I'm not sure.

One intermediate point which you made is, of course, you could also try to do some kind of a dose response in the sense where the level of the surrogate is correlated with the clinical response. If there is a treatment effect and if the surrogate is doing what it should be doing, then there should be some kind of a correlation between the surrogate level, if you wish and perhaps the change in volume or the lack of change in volume, with the change in cognition, say. That would be simpler trial probably, then some kind of a randomized withdrawal which has practical as well as ethical problems, maybe.

But that's what I would start with. And it's pretty humdrum and pretty traditional, but that's where I think I would put it at this time.

The other thing we were talking about trials of a year. These are very hard trials to do with older patients. You know, there's death dropout. The patients that are deteriorating the most rapidly are most likely to withdraw. You really are going to have a hard time with any kind of a trial to prove the efficacy of a surrogate, if you wish, if you're going to take that long a time, yet that's what the people here around the table suggest.

The other thing, the final thing I'll mention is that how tolerable are these procedures, again, for the reasonably advanced Alzheimer's patient in terms of agitation in terms of where they're at? You know, somebody saying an MRI takes an hour. That, I assume, is not just sitting under the instrument for an hour. But what kind of a time line are we talking about and are we bordering on patient abuse just to satisfy a clinical question? I just raise that as an issue.

DR. FOX: I just wonder whether I could both make comment on the tolerability issue and perhaps also since my comments on what I felt in terms

of sustainability or effect would have on terms of reasonably likelihood, if I may make comment?

Firstly, on terms of patient's tolerability, I can really only speak for volumetric MRI and sequence takes that we use for the images you saw takes 10 minutes. We're nearly completing a study of 50 Alzheimer's mild to moderate, meaning mini-mental of 19 where they have 9 scans over a year. We've had 5 percent dropout and 5 percent missed scans. And one of the people who had a mild cortical infarction was keen to come back and complete the rest of his scans.

So, the care and attention of the investigators, the people find very supportive in this extremely distressive disease. And having MRIs does not, in my opinion, have major contribution. Yes, some people will be claustrophobic. In my experience we have a higher number of claustrophobics in the controls than in the Alzheimer's patients. And, yes, some people can't stay still. So that's, as far as I'm concerned, about that issue. And people are desperate for that care and regular attention.

And as far as the sustained effect is concerned, I was trying to make two points that in my very brief presentation. One, yes, withdrawal I think adds support. But the other point I was trying to make is we set up lots of hypothetical situations here. And I'd like to put one which makes the point about the sustainability, having watched my at-risk cohort see hippocampal atrophy progress inexorably and their whole brain, and then seen the clinical progression follow that. And it was always very compelling to me if that -- at least the natural history component.

But the sustainability, I have two parts. One withdrawal, which I think adds confidence, and the second is I think there is a different level of confidence one has if, for example, if I scanned you monthly for 6 or 12 months and saw that the rate at each of that time was being consistently reduced; I have a greater level of confidence in that and a greater likelihood that a reasonable person would think that was associated with clinical outcome than if I just had a first and a last scan.

That was the point I wanted to make. I

think both add in my opinion a level of reasonableness, if that's a word.

DR. GRUNDMAN: I basically agree with that. I just think that so in terms of Dr. Katz' question about what would you include in a trial to, I guess, try to make it more likely that you would accept a surrogate as reasonably likely that it's going to improve the clinical outcome, is that the question?

DR. KATZ: No. Again, the use to which I think we're mostly talking about these imaging modalities would be put would be to support a claim for an effective drug on a progression, on the underlying progression of the illness.

A number of you said that we're not ready to come to that conclusion on the basis of an effect solely on the surrogate. So what I'm asking is, is there a trial design that we could do now that would support an effect, a claim for an effect on progression? Would it combine the imaging plus a clinical? Would it combine imaging plus clinical plus a withdrawn phase? I'm just trying to get a sense of

what, if any, design you think we're ready to employ to address the question of an effect on the underlying progression of the disease. It doesn't even have to include a surrogate.

DR. GRUNDMAN: Okay. So I would say two things.

One, if you were to do a clinical trial -- first of all, you know, again it always depends on the clinical group that's involved in the trial. Because, you know, your sample sizes are going to be smaller and the trials are going to be a shorter if you're doing an AD trial than if you're doing an MCI trial, or if you're doing some sort of a prevention trial.

So if we're just talking about AD trials here, I think that you could do a clinical trial with clinical outcome measures that we're used to, CDR, CGIC, ADAS Cog, classical outcome measures and measure MRI. And if they are both consistent with one another, then I think that that would support, you know, a disease progression modification type claim. I think it would support that. Obviously, it doesn't prove it, but it would at least be consistent with that notion

if you saw, you know, half the rate of decline on the clinical and cognitive measures and you saw some diminution in the rate of decline on the MRI atrophy; that would at least be consistent with that conclusion.

DR. KATZ: Any minimum duration?

DR. GRUNDMAN: For the trial?

DR. KATZ: Yes.

DR. GRUNDMAN: You know, I think it sort of depends on how many people you have in your trial. Because you can show an effect with a larger number of people over a shorter period of time.

DR. KATZ: Right. But -- perhaps. But what I'm asking you is there any minimum duration below which you would say well, I see they both go in the right direction, but that doesn't prove to me that there's an effect?

DR. GRUNDMAN: I think probably -- you know, I would say just practically for the clinical measures you might need a year to show that with several 100 people in each group.

I think for the MR measures, I'm not sure

how long it would take because I'm not sure it's quite as well worked out.

But I think one other point is that if MR measures were conducted in the context of a clinical trial and you were collecting them in a sequential fashion, say every 6 months, and you did, say, an 18 month or 2 year study and you found that the effect on the MRI occurs at, say, 6 months and the effect on the clinical occurred at 18 months, that would help give me confidence that if they did another trial with that agent, that if you found an effect over that short of period of time, that might support an accelerated approval.

DR. PROVENZALE: Comment?

CHAIRPERSON KAWAS: Yes. I'm not sure it's an ideal situation, but I would suggest that a withdrawal design probably needs to be incorporated to convincingly make the case for disease modifying.

DR. KATZ: Including an imaging surrogate or just clinical, or just imaging?

DR. WOLINSKY: Yes. I don't think it's very practical to think about withdrawal designs. I'd

rather be an optimist. I'd rather believe that the treatments are going to work, and I'd rather deal with the issues of the so called randomized start or delayed start of therapy.

Well, when you start using something like atrophy as the endpoint of measure if these curves never catch up, that's telling us something very important. And that's what we would expect to see. And certainly for some of the studies we've done in MS where we've had randomized starts of a sort, these are exactly the kind of curves we see.

DR. KATZ: No, that's fine. Again, I'm just asking what the elements of such a trial would be. We can refer to them sort of generically as a switching maneuver or whatever you want -- let's say randomized start.

DR. WOLINSKY: Yes, but one loses patients and the other one keeps them.

DR. KATZ: Well, no, fine. Again, I'm just trying to understand whether or not you need some sort of a phase like that at all or whether just a simple parallel study which shows a correlation at

some point down the road between clinical and surrogate is sufficient. I'm just trying to get a sense of what people think.

DR. PROVENZALE: Comment. With regard to design of the length of the study, that is I think largely governed by what change you're hoping to see, what would be statistically significant. I mean, you know, let's say going back to what Dr. Jack showed about the hippocampal volumes. You know, a statistician would basically have to decide, you know, would a difference between 1.5 percent decrease and 3.5 decrease; that's the annual rate of change, I believe, that he gave us. You know, would that be long enough or would you have to have 2 years at those different rates in order to see a statistically significant difference between the two? If you remember, the standard deviation there was relatively high for those rates of change.

So, I mean, I don't think that this is a question that we an answer without a calculator, basically.

DR. KATZ: Well, I take your point. But

it's true that depending upon what the treatment effects size is, and you know, and the rate of change you might need -- if you enrolled a lot of patients, you might be able to do a shorter study. But I'm trying to find out whether or not as part of the element of this theoretical study I'm trying to decide whether or not even if you could show an effect with 500 people at 3 months, for example, is that satisfying. The fact that you could show an effect doesn't necessarily mean that you would believe that that is an effect that is structural and would persist.

So I'm just trying to get a sense if people think, well even if I could show it at 3 months, it still wouldn't convince me it's a structural effect. I still want to see at least 6 months or at least 12 months.

I recognize that everything we're asking you today is hypothetical.

CHAIRPERSON KAWAS: Well, hypothetically, I would like at least 6 months. And I would like a combination of cognitive and imaging. And you take

the GCIC, substitute the imaging.

DR. GRUNDMAN: I would say you can take the CGIC, but I'd still like to see some sort of measure of clinical change as opposed to just simply cognitive change in the trial.

CHAIRPERSON KAWAS: Like what?

DR. GRUNDMAN: Like the CDR sum of boxes, some sort of measure of function.

DR. PROVENZALE: It depends on what. Are you talking about an AD study or an MCI study.

DR. GRUNDMAN: We're talking about an AD study.

DR. PROVENZALE: It totally depends on -- the problem with a CGIC, I think there is a problem with that because the longer the trial, the harder it becomes to remember what's going on at the baseline. So you do need some sort of functional severity measure which can be assessed serially over time which doesn't depend on the person's recollection of their baseline status.

CHAIRPERSON KAWAS: Any other takers for Dr. Katz' questions? I've never seen a committee so

quiet.

Dr. Love, how about the questions that you posed for us? Have we approached them in anyway helpful or could you like guide a little?

DR. LOVE: I think, yes, I think they've been helpful.

Obviously we are speculating at this point in time and looking for approaches to use in designing these trials and not just for Rusty's purpose of what are we going to do for a drug when it comes for approval. Because these questions are being asked now and these studies would need to be designed at this point in time, but also these studies may be useful to help establish the reasonable likelihood aspect of this.

So, there would be probably features that we want to think about based on things that you've all mentioned. Maybe if you're looking for clinical effect at 3 months, does that means that the functional imaging measure should be timed along with that if you think the functional imaging may change before the anatomic image does?

So, just those kinds of thoughts would need to go into the design of the trial. But I think we've heard a variety of comments and issues we'll need to continue to think about.

Yes?

DR. GRUNDMAN: Well, I was just going to say, you know, I think that the functional measures or the functional and the anatomic and the clinical measures should be done simultaneously. But I think they should be done serially over a period of time so that you can compare the simultaneous measures with each other, and then you can also compare the imaging measures with their ability to predict the ultimate outcome that you're looking for in the trial.

CHAIRPERSON KAWAS: Dr. Van Belle?

DR. VAN BELLE: Just one final comment. I was somewhat negative today, but I'm actually quite excited about imaging, and I do think it's a very useful technique. Some of my best friends do imaging.

I just think that should it be done in a proper scientific context. And I think the rules for that are actually fairly straightforward and are known

to the FDA as well as to the industry. It's just a matter of doing it.

CHAIRPERSON KAWAS: I'll ditto everything Dr. Van Belle just said.

And also in follow-up to your comment, Dr. Love, I can easily imagine a surrogate marker like one of these images being positive long before the clinical outcome. It just doesn't have to be, wouldn't necessarily be. I mean, I can se how both of them in some cases depending on what the drug is doing, could become positive together. But I can easily see, I mean the imaging becoming positive before the clinical, whereas the opposite is quite hard to imagine.

DR. LOVE: Right. But that's the type of information that would be useful in the long run to determine that this is truly a surrogate or at least reasonably --

CHAIRPERSON KAWAS: Right.

DR. WOLINSKY: Okay. I would sorry about that a little bit just from experience in a different field. Because some of the metrics that we measure in

MS seem to be dependent upon events that may occur a year or two earlier; that the dye is then cast so that if there's an effect of the drug it may take the third or fourth year until you begin to see the effect of that drug.

So I'm not sure until you know, and I don't know about MS, but I do get the feeling that we may know a little bit more about that than Alzheimer's, I'm not sure that we actually can make these predictions. And, therefore, you have to be very careful that you cast long enough nets for your data.

DR. LOVE: And that probably goes to one of Rusty's questions, how long -- how long should the studies be. There's the short term how long --

DR. WOLINSKY: But assuming that we see in an Alzheimer's Disease is actually locked in some kind of a way, and I'm not sure that it is, it looks to me that the data that Dr. Fox had presented and some of the other data would say that these studies have to be as a minimum to rigorously detect atrophy about a year and probably 2 years. And that means that if you're doing at the late start, you're into a fairly long

trial.

CHAIRPERSON KAWAS: Well, does anyone else from the Committee have any comments, issues they want to bring up? Any discussions? Any of our invited speakers? Dr. Doraiswamy?

DR. DORAISWAMY: One of the comments that was made was the time course of whether your clinical outcome would turn out to be positive before the imaging outcome or vice versa. I mean, we know from our clinical trials already that the ADAS Cog becomes positive around 6 weeks in many of the drug trials. I'm not sure that most people expect a hippocampal or a brain atrophy volume to become positive at 6 weeks. So in most trials the brain volume changes are probably going to occur after the clinical outcome sort of changes. That depends on what outcomes you're looking at. Obviously, if you're looking at survival or nursing home placement, etcetera, then probably your imaging outcomes will predict those. But certainly not the ADAS Cog outcome, because we already know that from the clinical trials.

So, I just thought I'd throw that out.

CHAIRPERSON KAWAS: Well, I just want to comment that we know that from trials that were symptomatic trials. But if we're talking about disease modifying trials, which I think is actually one of the interests, and we're talking about -- then that's where I think we're going to see the imaging positive before we're able to detect a difference in the rate of cognitive decline. But in symptomatic trials, absolutely. I mean, you fix the symptoms before you fix the underlying disease in terms of time always.

Yeah. Dr. Fox?

DR. FOX: I was just going to agree with what you said in that the power calculations for disease modification suggest that you'd be likely to see -- if you had a purely disease modification effect, you'd be likely to see the effect on your surrogate before your clinical. And it's quite probable that you have both the symptom -- you may well have a symptomatic and a disease modification or one without the other. It's possible.

DR. DORAISWAMY: And you may never see a cognitive effect. Because in the Vitamin E trial, for

example, Vitamin E did not have any effects on the ADAS Cog at all. And it's possible, I mean no one's looked at brain changes in relation to Vitamin E therapy, but it's hypothetically also possible as you and some others indicated that you could get a disease modifying agent that effects brain structure without effecting cognition at all. That's a theoretical possibility.

CHAIRPERSON KAWAS: I mean, I think the disease modifying trial is really the important thing in the field as well as the discussion that's happening today. I don't think most of us are that worried about whether or not these are useful for symptomatic -- drugs that give a symptomatic effect only.

Dr. Katz?

DR. KATZ: Just to respond to what Dr. Doraiswamy said, if you had a drug that effected a surrogate but had never had an effect on cognition, I'm not sure what you'd have. In fact, I think it would suggest that the surrogate's a failed surrogate.

CHAIRPERSON KAWAS: Does anyone disagree

with that statement? I mean, I think that's exactly right. The only issue is that if it takes a longer time to show disease modification on a cognitive outcome, but the assumption would be that eventually you would be able to demonstrate it if the surrogate was a valid one.

DR. GRUNDMAN: I was just going to make one point about Dr. Katz' question about whether or not to do randomized start designs with imaging and so forth. And what was the point I was going to make?

I think that --

CHAIRPERSON KAWAS: What's the volume of your --

DR. GRUNDMAN: Oh, the point was that, you know, a lot of these drugs could have both symptomatic and disease modifying effects. And so the waters get sort of muddied when you do those designs and the curves go back but not completely. So then do you power your study to show that residual difference or, you know, I think those types of designs become pretty complex when you have to deal with them in reality and not just, you know, in a theoretical construct where

the differences are maintained at the same level that they were when the randomized portion of the trial ended.

DR. KATZ: Well, I think things have the potential to get extremely murky if you had a drug that had an effect on the surrogate -- I won't say in progression yet, but on the surrogate and also had a symptomatic effect. Because if you did a short term study with a drug like that, you might see an effect on the surrogate and you'd see your clinical effect, and you want to conclude that this must have an effect on progression because there's a correlation. But, in fact, it may have two actions. And the effect on the surrogate may actually translate into absolutely nothing clinically.

So, it's very complicated. Although I suspect in a case like that, if you did a randomized withdrawal and you maintained -- I don't know. But if you maintained a difference on the surrogate but perhaps the clinical outcome went back to where it was, you might argue there still was some effect on the underlying structure. What that meant clinically

you still wouldn't know.

CHAIRPERSON KAWAS: Okay. Well, it's been a very long and interesting day. And I want to thank all of the members of the Committee, all of the invited speakers, the FDA, the audience, and I think Dr. Katz has some comments for us.

DR. KATZ: Well, I just also want to thank everybody. It's been a long day, an interesting. It's a lot of complicated issues.

I appreciate our invited speakers coming, the imaging consultants and the neuro committee.

And I neglected to thank one person when I spoke earlier this morning who really is largely responsible for the meeting at all, that's Dr. Ranji Mani who is a senior reviewer in the neurology group in the division who identified the experts and really wrote our briefing documents, and really put together the whole meeting. So he deserves our great thanks.

(Whereupon, at 4:54 p.m. the meeting was adjourned.)

1