Neuroimaging in the Diagnosis of
Alzheimer's Disease and Dementia

April 5, 2004

Bethesda, Maryland

The Neuroscience and Neuropsychology of Aging Program, National Institute on Aging (NIA), DHHS

and

The Centers for Medicare and Medicaid Services (CMS), DHHS

Welcome and Introduction

Neil Buckholtz, Ph.D., National Institute on Aging

I'll be your timekeeper. Susan Molchan, my colleague, will be introducing me at the end. But, I want to right now introduce Susan Molchan, also from the National Institute on Aging, to make some comments, and she'll be introducing Dr. Steve Phurrough from CMS.

Susan Molchan, M.D., National Institute on Aging

Thanks for coming, everybody, this morning, especially this morning. We're all supposed to be springing ahead. I'm just going to give you a little background on how this meeting came about and why it came about. It arose, from what I understand,

essentially from a request from CMS, or from Tommy Thompson to CMS, to have an expert panel meet to provide them with information on the question of the value of PET in the early diagnosis of Alzheimer's disease. CMS then initiated a meeting with NIA last August to discuss the formation of such an expert group. And, after meeting with us in August, we all decided to hold a meeting in the Spring, this meeting, to address specific questions that CMS has on the diagnosis of dementia and Alzheimer's disease, and what PET adds to that diagnosis. We'll also be getting information on the role of other imaging modalities, on MRI and SPECT, and what role they play in early diagnosis. Lots of research, of course, has been done in MCI and Alzheimer's disease, looking at various regions, especially in MRI.

Everybody would agree, I think, that we've assembled a stellar group here today, with top scientists in Alzheimer's disease imaging and public health research, as well as leaders from the two advocacy organizations, the Alzheimer's Association and the American Association of Geriatric Psychiatry. So, I think our CMS colleagues and everybody attending here today will leave with an up-to-date overview on the state of Alzheimer's diagnosis and on the issues to consider in evaluating the use of imaging technologies in the diagnosis of Alzheimer's disease.

So, I'm going to introduce my colleague from CMS, Steve Phurrough, and he'll give an overview of the perspective from CMS.

Steve Phurrough, MD, MPH, The Centers for Medicare and Medicaid Services

Thank you, Susan. Again, I'm Steve Phurrough. I'm the Director of the Coverage and Analysis Group at CMS. It's our job in this particular group at CMS to make decisions about the kinds of things that we're going to pay for, and we have been doing this in a very formal, evidence-based fashion for about a decade or so. And, about half of that period of time, for close to five years, we have been addressing some issues around Alzheimer's disease.

We first addressed it formally in what we call our National Coverage Decision Process. That's our evidence-based process for looking at the particular technologies or services that we are being asked to provide reimbursement for. We produced one of our decisions in December of 2000 where, after consultation with a number of folks, we first chose to not provide reimbursement for PET scans in Alzheimer's disease. And, the recommendation at that particular time was that we undergo some reviews with our advisory panel; our organization has a Coverage Advisory Panel, who then provided us some additional suggestions and guidelines, and we produced, with a number of people's input about this time last year, a fairly extensive discussion of whether we should broadly cover the use of PET scans in the diagnosis of Alzheimer's disease. At that particular time, we continued what we called our non-coverage. However, at that particular time, the Secretary thought it important that the information gathering continue, and he directed us to both look at an expert panel to address the issue, as well as a demonstration project. And, we immediately came to NIA to discuss with them the potential of assisting us in the expert panel.

We want to express our thanks and appreciation to NIA for putting this together. We think this is a superb agenda, a superb group of people today to help us answer the questions of the appropriate imaging techniques for Alzheimer's disease, what's the state of the art, and perhaps more importantly, what else do we need to know as we look at issues around the diagnosis of Alzheimer's disease.

Even since that time, since we first began talking with NIA late last summer, we do have another request now for more narrow coverage of PET scanning in Alzheimer's disease. And you have, in your little blue folder, some questions that we are interested in answering, some fairly specific questions around what's the appropriate use of imaging in the diagnosis of Alzheimer's. And, so, we hope to hear some of that discussion today so that as we, once again, look at the issue, we are able to use your knowledge in our attempts to most appropriately provide for our beneficiary population.

We obviously have, in our Medicare population, large numbers and ever-growing numbers of patients with Alzheimer's disease. We certainly are interested in the appropriate treatments and appropriate diagnostic technologies and look forward to the discussions today to help us achieve that benefit. Thank you all again for coming and, gain, special thanks to NIA for helping us with that.

Neil Buckholtz, Ph.D., National Institute on Aging

Thank you, Steve and Susan. I have just a few announcements to make before we actually get started. Actually, we're running a little early, which is good, so that'll give

4

us some more time for the discussion section. Each of the presentations this morning will be 20 minutes. We've decided to hold the questions until the discussion sections. I think that's probably going to be a little more efficient. So, if you have questions, please write them down and, then, we'll have a very open discussion section for each of the sections that we have. Because some of the presenters had to come in at different times and leave at different times, we've set it up so that, for example, John Trojanowski, who really should be in the first section, is going to be a little later. And, so, the way we've broken it up is that there will be actually two discussion sections after the break, one on the first set of presentations and then another one on the second set of PET presentations. So, it's a little funny when you look at the actual agenda, but that's why, because we have some speakers who are having to come in a little later and leave a little earlier.

I'm going to be the timekeeper for the presentations. We've found that this is a very high-tech solution to keeping people on time. I have various cards here – (laughing) – which I'll put up and make sure that people can see. And, this is the one, "We have no more time." So, I will come up and kind of gently nudge you off and have the next speaker come up. So, I'll be sitting over there, so for the speakers, please keep an eye on the time.

We will have a working lunch, because we're on a very tight schedule, and we need to get people out absolutely by 5:00. And, as I say, some people have to leave a little earlier. The Federal employees have to pay for their lunches and, so, if you will

please check with Vicki Hill out here, if the Federal people want to have lunch. Everybody else is covered, so you don't have to worry about it.

So, with that, as I say, we're running a little early, which is okay. I would like to introduce the first speaker, Ron Peterson from the Mayo Clinic, who's going to be talking about accuracy of clinical diagnosis of Alzheimer's disease and differential diagnosis. Yes? (Inaudible/off microphone) Oh, right. The speakers need to mention any conflicts of interest that they have before they begin speaking. Thank you.

Ron Peterson, M.D., Ph.D., Mayo Clinic

Thanks, Neil. While we're getting loaded up here, I'll just say that I do consult for Elan Pharmaceuticals and, actually, probably independent of this activity, but might be relevant, I'm forging a new consulting relationship with GE, actually, as we speak, get some research funding from Pfizer and ETSI through the Alzheimer's Disease Cooperative Study, and that's it. I'm not on any speaker's bureaus.

My task this morning is to sort of set the groundwork or the foundation for interpreting some of the subsequent information. In other words, if we're talking about making the diagnosis of Alzheimer's disease, at what point do we start? That is, where do you start introducing various technological advances and what's the baseline? So, the baseline, I think most people feel, is with a clinical diagnosis. And the point being, then, who makes that diagnosis and how can we augment that? And, I think most people feel, in the fields, that this should be made by a clinician. Clinician could be family

practitioner, internist, neurologist, psychiatrist, somewhere along the way. The clinician then is charged with taking a history from the patient, from family members, doing an examination, including a mental status examination, plus/minus, say a neuropsychological testing, and then rendering a diagnosis.

So, what I'm going to address this morning is where do we pick up from that point? Where are we when we get done with the clinical diagnosis? Most of the data that I'll be discussing comes from a practice parameter paper done by the American Academy of Neurology a couple of years ago. There have been a couple of papers since then, but I think the basic message that was forged in this paper still holds true. As you know, the process of generating a document based on evidence-based medicine is quite long and tedious: Forming the clinical question, doing a thorough literature research, as well as possible supplementing that by the panel of experts from their own personal history in dealing with the disease, evaluating the quality of each of the articles with regard to class one, class two, class three types of data, what are the recommendations that come out of that process and, then, send the paper out for review by a variety of expert panels. So, this is a very painstaking process but generally generates a document, then, that is useful for the field. Finally, after the advice of the various panels comes back, the final document is then created.

The primary question, of course, is what is the accuracy of the clinical diagnosis here with regard to if you take the pathologic standard as the gold standard. how accurate are you when you say, yes, the disease is present versus whether, in fact, the disease is

present, generating sensitivity and specificity-type figures that we're all fairly familiar with.

The clinical questions addressed in this paper, then, regard are the current clinical criteria for the diagnosis of dementia and Alzheimer's disease reliable, and are they sufficiently accurate, that is, are they in fact valid? They also address laboratory tests. We're going to get into that with regard to PET later in the day, so I won't address that specifically right now. As we're all familiar, the essential criteria for dementia, then, are that the person has an impairment in cognitive function, usually involving memory and other cognitive domains, and these disturbances are of sufficient severity to compromise the person's daily functioning, hence they are demented, and there's not another explanation for that delirium, drug toxicity, etc. So, these are the essential criteria that all clinicians have in their heads when they see the patient who comes in with some type of a cognitive complaint and then is trying to determine whether, in fact, the person meets clinical criteria for dementia.

Based on that, the panel then reviewed the extensive literature in the field and came to the conclusion that, yes, the criteria such as DSM3R or like published criteria for dementia are, in fact, reliable and valid. And, this is based on a variety of class one studies, indicating that the kappa varies from 0.5, 0.7, 0.9, 0.5, 0.7, sort of marginal, 0.9 quite good. But, in fact, the clinical diagnosis of dementia in the hands of experienced clinicians is actually quite accurate.

With regard to validity, there's one class one study that, in fact, indicated that the diagnosis was quite accurate, as well. So, based on this, the panel recommended that DSM3R criteria or similar criteria are, in fact, reliable and should be used by clinicians. So, in fact, the basic clinical standard is pretty good. When you operationalize those criteria, as clinicians, you tend to be right more often than not. So, the standard is quite high when you're starting to augment that with any type of ancillary test.

But, what about moving on to probable Alzheimer's disease, then, going from dementia to Alzheimer's disease? Again, how do you make that diagnosis? Dementia is established by a clinical examination, documented mental status tests, perhaps with neuropsychological tests. Again, you have to have a deficit in two or more areas of cognition, usually involving a progressive disorder of memory and other cognitive functions, in the absence of another explanation. So, we're not dealing with brain tumors now; we're not dealing with infections but, in fact, this is thought to be due to a degenerative process, no disturbances in concentration, consciousness, and in the appropriate age range. If people meet these criteria, then, the likelihood is that they have probably Alzheimer's disease. Based on these, the reliability is again pretty good, 0.5 to 0.7, looking at a variety of class one and class two studies, but a fair amount of data out there indicating that, in fact again, clinicians are pretty good at making this diagnosis.

When you put up the standard against pathology, again if we take the gold standard as being neuropathologic confirmation, for probable AD, there are 13 studies, three class one studies, 10 class two studies. Sensitivity is pretty good, 81% with range

from about 50 to 100%, depending on the individual study. The specificity is okay but not quite as high, and this is something that we may get into a little bit later, in fact, that the criteria for Alzheimer's disease are a bit broad and can be encompassing of other dementing illnesses, as well. With regard to possible Alzheimer's disease, four studies indicate again the sensitivity can be quite high now, because possible AD, the diagnostic criteria here are a bit more broad, more encompassing, but the specificity drops to only 48%, with the range being 32 to 61.

How common is Alzheimer's disease? Well, if you look at a variety of studies out there involving pathology, you find that pure AD ranges, 55, 61, 38%, but AD plus other comorbidities can be quite high. So, in fact, the condition is out there in a prevalent fashion. It's there, and this influences again the accuracy of our clinical diagnosis. I'll come back to this issue of what is the prior probability that somebody has Alzheimer's disease when they walk in your door, and that can influence the outcome of these studies.

If you map sensitivities and specificities on a variety of studies that have been done, you see that they do vary quite a bit, from a low here in the {Jobe's Study} of under 50% sensitivity but very high specificity, to an early study from Washington University with very high sensitivity and reasonably high specificity, as well. So, they do vary, but I think if you look at the sensitivity here, we're in the 80 to 90% range. Specificity then drops 60 – 60+% is probably a reasonable representation. But, now, if you superimpose upon these studies the probability that somebody has Alzheimer's disease when they walk in the door of these institutions, I think you get another

perspective on this, that is, here's the likelihood and, again, this reflects how people get

into studies, how they get in particular studies that lead to autopsies. We all know that

while autopsy may be the "gold standard" in a sense in the field, only a subset of people

come to autopsies. So, you have one selection factor operating by who gets in the front

door of the clinic and, then, once they're in the clinic, what subset goes to autopsy? So,

you have to take those factors into consideration.

So, here, you can see in the Jobe study that less than half the people ended up with

the diagnosis of Alzheimer's disease. So, they're dealing with a somewhat different

clinical situation than, say, the Wash. U. study up here, where 90+% of the people who

walk in the door get the diagnosis of Alzheimer's disease, hence the fact that the figures

are going to be quite high and, down here, they're quite variable. So, in the field, you're

probably looking at a clinical diagnosis of Alzheimer's disease in the 60 to 70% range

people walking in the door and that, of course, then influences your subsequent

sensitivity and specificity figure.

So, based on these findings and all of these considerations, then, the American

Academy of Neurology issued the guideline that probable Alzheimer's disease, clinical

diagnosis, is lacking in sensitivity for atypical pathologic AD variance, possible AD

clinical diagnosis, lack specificity as it captures many cases with, say, FTD and vascular

features. So, this is something that we can, in fact, address later on. Can we improve

upon these sensitivity and specificity figures with any ancillary imaging procedures?

Now, expanding then to the differential diagnosis issue, again, if you look at the diagnosis of dementia and you look at the subtypes, putting a bunch of studies together, many on the earlier slides, you can come up with a figure that perhaps half or slightly more than half of all dementia diagnoses end up being relatively pure Alzheimer's disease. If you add other concombinant factors, like vascular factors, you may pick up another 10%, so Alzheimer's disease with vascular factors picks up 10. Pure vascular dementia, relatively uncommon but not trivial, if you add in Lewy bodies these days, you're going to pick up another 10% or so of AD plus Lewy bodies, and pure Lewy body disease, relatively rare again. Then, you pick up the frontotemporal dementias, and these numbers bounce around in terms of where are we, eight, 10, 12% of all dementias that might be frontotemporal and, then, the grab bag of all the other things that we see. But, here, if you put the AD plus its cousins together, you're going to get over 70% of all diagnoses have an AD or AD-like component.

So, now, we're asking the question: Can imaging or any other biomarker type test help differentiate among these. The primary challenge for PET, then, where PET might be useful, is can it do a distinction between Alzheimer's disease and, say, frontotemporal dementia? What do we mean by "frontotemporal dementia"? Various criteria are out there. Most of them sort of distill down to a group of disorders with primarily frontal features, with behavioral changes, personality changes, loss of insight, or the temporal variance of these disorders, with language changes, either fluent, non-fluent, or semantic dementias, something like that. The clinical pathologic correlation in frontotemporal

dementia is much more challenging, fewer studies out there. One of the studies that was in the practice parameter paper by Varma indicated, though, and this is something that we'll probably get into discussing later on today, that while 26 subjects had FTD changes neuropathologically, 20 of them would have also met clinical criteria for Alzheimer's disease. So, the Alzheimer's disease clinical spectrum might include FTD if you apply it somewhat loosely.

At the American Academy of Neurology later this month, Dave {Notman}'s going to present some data from our group, indicating that of about 27 neuropathologic FTD diagnoses, the sensitivity was 81%; the specificity's quite high, but that's just because there are so many non-AD cases that makes the performance here look better than what it might normally be. But, in general, if you put together a combination of clinical, neuropsychological and imaging techniques, generally speaking, a clinician can make the diagnosis of frontotemporal dementia fairly accurately. The clinical criteria, then, range from insensitivity from 63 to 73%, while the specificity is in a rather high range again. But, this is due to the fact that there are so many non-FTD dementias out there that the denominator is quite large.

Now, the other area in which PET scanning has been suggested that it might be helpful is in the very, very early diagnosis. I won't spend too much time here, because there aren't a lot of data, but this is felt to be a clinical state that's intermediate between the cognitive changes of normal aging and dementia. Criteria are out there but are under discussion, somewhat variable, and probably the most precise criteria deal with the

pre-Alzheimer's disease form of mild cognitive impairment, and these are reasonably accurate for predicting who's going to progress to the clinical and neuropathologic diagnosis of AD. With regard to path studies, very few out there; probably the best one is from the Religious Orders Study where the conclusion was that, now, these are people who died while they had the clinical diagnosis of mild cognitive impairment. In the Religious Orders Study, they used a diagnosis that involved multiple domain memory; plus, other cognitive domains are involved. When people died in that clinical state, about 60% of them in general had AD-type changes at that point in time. A study from our place indicated that something fewer, then, 50% had AD at that time, but another study indicated, in following the people longitudinally after they progressed from clinical MCI to clinical dementia, now about three-quarters of them ended up having Alzheimer changes. Study from Wash. U. on early Alzheimer's disease, again following people who came into the system with a CDR of 0.5, following them clinically, seeing what their pathology was and the outcome, 84% of them developed AD-type changes. So, it looks like, in these intermediate stages, if people meet criteria and are followed longitudinally, three-quarters to four-fifths or so will, in fact, have Alzheimer's-type changes. But, again, these are preliminary studies, and there are not a lot of data out there.

So, let me conclude that I think the clinical diagnosis of dementia, the clinical diagnosis of Alzheimer's disease is pretty accurate. Most clinicians make this diagnosis correctly, probably in the 80-85% range. So, the challenge now is to improve upon that particular figure.

The FTD clinical diagnosis is moderately sensitive, quite specific but, again, the current criteria that are out there have a fair amount of overlap between Alzheimer's disease and FTD. If you take the Alzheimer's disease clinical criteria, and you take a patient with frontotemporal dementia, that patient will probably also meet criteria for Alzheimer's disease. So, in that sense, they're not mutually exclusive, and that's something that I think we have to address and deal with.

MCI, it's pretty early to tell. It's likely that the amnestic form of MCI goes on to develop Alzheimer's disease, but there really are no data out there on imaging utility in the early stages of this with longitudinal follow-up.

So, I think this sets the stage, then, for where we are clinically, and we can take questions later on. Thank you.

Neil Buckholtz, Ph.D., National Institute on Aging

Thank you, Ron. The next presentation will be from Denis Evans at Rush Presbyterian St. Luke's in Chicago, and Denis will be talking about prevalence and incidence of AD, other dementias and projections for the future.

Denis Evans, M.D., Rush Institute for Healthy Aging

Well, thanks very much, Neil and all the people. It's a pleasure to be here. I'm going to talk about work that I played a relatively small part of here, and most of the work was done by Liesi Hebert and Paul Scherr. Liesi's in our group, while Paul is at the

Centers for Disease Control. Julia Bienias was the statistician, and my colleague, David

Bennett, was a guiding force in the neurological diagnoses. And, I wish to talk a little bit

about what's here first on the slide, the prevalence and incidence of Alzheimer's disease

and just point out some practical problems in counting people who have Alzheimer's

disease.

And, most of the talk, I want to devote to some estimates of what's going to

happen in the future. Are we going to have more people with Alzheimer's disease and, if

so, why? And, proceeding to that first part, that is to actually talk about challenges and

measures of occurrence of Alzheimer's disease. There are really about four of them,

okay. The first is that, like some common chronic diseases and unlike others,

Alzheimer's disease is typically not well-detected in the routine delivery of medical care.

If you're the CEO of a Fortune 500 company, and you start to act as though you're

demented, the likelihood is that this will be picked up, especially if you're around the

age of 60 to 65 and still occupationally active. However, if you're an 85-year-old black

woman who has outlived her husband and most of her near relatives, who lives alone far

from any sources of medical care, the likelihood that you will be detected as having

Alzheimer's disease is relatively small. So, medical care only detects a fraction of those

people who have Alzheimer's disease, and this fraction is typically biased by the factors

that determine access to medical care for this and for other conditions.

A second difficulty is, like many common chronic diseases, Alzheimer's disease

comes on by minute degrees. It isn't like you went to bed one night okay, and you awoke

the next morning with Alzheimer's disease. That of course may occur in a rare case, but

it is an extremely atypical path to take. Most of the time, whether one has Alzheimer's

disease or not, the cut point between normality and having the disease varies, and it

varies according to judgment. I might judge that a substantial percentage of people who

are in the middle actually have Alzheimer's disease and someone of a considerably more

conservative pace might judge that none of them do. Obviously, we'll come up with

different figures for the occurrence of the disease. We often use, we all use brief

cognitive tests to determine somebody's cognition. That, of course, is at the very heart

of the disease. But, sometimes, we use these tests for screening a population, that is,

someone has to have poor performance on subcognitive tests before they're judged

suitable for full evaluation for whether they have dementia or not. The trouble is that

cognitive tests, in general, are relatively insensitive ways to screen the population for

Alzheimer's disease, so that if one applies such a screening procedure, one is really

giving you the prevalence of Alzheimer's disease among those who have poor

performance on the cognitive tests and ignoring the relatively small prevalence of

Alzheimer's disease among the rest of the population. The trouble, of course, is that that

small prevalence of disease is applied to a very large fraction of the population, since the

usual reason we do cognitive tests or a screening test is to reduce the cost and the burden

of doing all these clinical evaluations. So, in essence, one might only really evaluate the

15 poorest performing percent of the population fully, clinically, until one is ignoring the

prevalence of Alzheimer's disease in the remaining 85% of the population.

And, a final problem just to regard the most common cause of the dementia in most studies other than Alzheimer's disease, vascular dementia, that's a particular problem, as Ron has alluded to, in criteria to distinguish other causes of dementia from Alzheimer's disease. All of these can cause prevalence assessments to vary substantially. Now, in just a moment, I'm going to present projections of Alzheimer's disease occurrence over the next 50 years. And, such projections, as we'll see in a moment, have a structure of going from a source study and then projections of the population, usually from other established sources like the U.S. Census.

And, the study I'll talk about here is the source study for the projections I'll be presenting but also, of course, it is a study in its own right for estimating the prevalence and incidence of Alzheimer's disease. In this study, we really considered as our population of interest the population of a geographically defined urban biracial community, actually, a population that is composed of residents of a geographically defined community of Chicago. Alzheimer's disease was measured uniformly in these people, independently of its ascertainment by medical care, our criteria detected relatively mild disease, as relatively severe disease, and persons were selected from this population for evaluation by stratified random sampling, rather than by screening. And, this slide merely shows the incidence, not the prevalence, of Alzheimer's disease, found in the source study, according to age groups. And, the only real point I wish to make with this slide is to note the dramatic increase with age, okay, ranging from 1.45% of all persons in the 65-74-year old age group, a little or quite a bit over 4% of those people 75-

84, and a startling 9.11% of people who are over the age of 85. These are annual

incidence figures, so you can see if you observed this population for a considerable time,

you would observe a considerable cumulative incidence of Alzheimer's disease.

And this, as I said, was the source study for our study of projections of the

occurrence of Alzheimer's disease in the U.S. population and the estimates of population

growth over the next 50 years were taken from the U.S. Census, and we considered the

effects of age, changing survival in the over-65 age groups, gender, education, and race

when we made these projections. So, all of these things, as one would think, are going to

change over the next 50 years, so we took all of them into consideration.

One technicality here that actually is an important one, we used incidence estimate

from the source studied and converted them to prevalence estimates in the projection

figures. The use of incidence estimates allows us to take better account of changes in

survival over the next 50 years. And, the major factors that affect the projections over

the next 50 years are aging of the population and survival of persons with dementia.

Gender, education have a much lesser effect on these. So, as an accrued estimate, one

can say that these projections are dominated by the effects of age and of survival of the

population.

A word about survival: We estimated that the general increase in survival which

is taking place in the older age groups in the population would affect persons with

Alzheimer's disease, too. However, their survival is not the same as those without

Alzheimer's disease; we estimate that that, from a previous study, that their mortality is approximately doubled. So, that is, the general increase in survival would not affect them, while it would affect both them and the people without Alzheimer's disease, it would not affect them uniformly, but in the ratio of 2.14 to 1. That is, they would benefit from the general increase in survival about half as much as those people who do not have Alzheimer's disease.

Now, because the projections that I'll present are going to be dominated so much by population aging, I want to present some measures of just how much population aging is affecting us. And, I wanted to do this in the form of what are known to most of you as "population pyramids." These show men on the left, women on the right, and they go by graded age groups, from the age of birth up to the age of 85. This was the U.S. population in 1950, and you can see a relatively small number of people in the top bar, 85+. I'll go through these very quickly, because it's really the visual impression that I wish to give you rather than boring you with a recount of what's going to happen to the population or what has happened to it. And, you can notice in 1970, the number of people over 85 has increased a bit, more in the women than in the men. And, by the year 1990, it's increased markedly and quite a bit more dramatically among the women than in the men. By the year 2010, you can see it's even more dramatically increased; it bears no resemblance at all to the structure we saw in the 1950 age pyramid. By 2030, the increases become even more dramatic; there's an enormous number of people over the age of 85 and, by 2050, it bears no resemblance at all to that structure that looked a bit

like the Empire State Building in 1950. So, over the 100 years between 1950 and 2050, the age structure of the U.S. population is going to change dramatically.

Just because one doesn't want to be terribly parochial, this aging of the population occurs not only in the United States but in all other developed countries of the world. This is the age pyramid for Japan; please notice that it's different over the one in the U.S., okay, in that the top age category is 100+, rather than 85+. This is 2000, this is 2025, and this is 2050, again, the dramatic changes there. This is Sweden, another country with a large old population already. The number of people over 85 in 2000 is very large; by 2025, it's striking. By 2050, it's very large indeed. And, to add a bit more drama, this is a country that we usually think of as a young country, much younger than us, of course. And, here, instead of extending out to approximately 16 million, those of you with sharp eyes can notice that the number of people in each age group actually extends out to 70 million, quite a bit larger than we have in the United States. In 2000, there's not too many people over the age of 85. In 2025, there's going to be a great many people in China over the age of 85. And, by 2050, there will be a very substantial number from this very large population of relatively young people right now.

So, all of that's to say that we can expect dramatic effects, because Alzheimer's disease is strongly related to age, and the age structure of the population is changing in such in such way that the number of older people is becoming much higher.

21

And, this is the number of people with Alzheimer's disease that we think will occur in the United States population between 2000 and 2050. We used three series by the United States Census to estimate this, three series that estimate population growth. What they really think will happen is estimated by the middle series, if you would, and to deal with the issue that projections of the population size, of course, involve a substantial amount of error when they're 50 years out. There is a high series of population growth and a low series, and one can see that by all estimates, by high, middle and low series, that the estimate of the number of people with Alzheimer's disease in the United States will approximately triple, we think, over the next 50 years. And, one can see here again, as you would expect from what I've just talked about before, the top graph is the total number of people from the previous slide, from the middle series of people with Alzheimer's disease; the range below it are the contributions from the three age groups. You can see the number of people 65 to 74 with the disease will remain approximately steady. The number of people 75 to 84 will increase quite a bit. The number of people 85+ is what's really driving this curve upwards.

And, because it's quite a bit unfair to give you just our group's estimates of what going to happen to Alzheimer's disease over the next 50 years, I would like to compare it with at least one other estimate that was done by Ron Brookmeyer and colleagues and published at the American General Public Health back in 1998. Ron and his colleagues used four separate source studies. A difficulty from my point of view would be that I think three of these studies probably underestimated the amount of Alzheimer's disease

in the older age groups currently, and those are the primary differences between those

projections and ours. The actual projection figures and the degree they'll be driven by

aging of the population, although not perfectly super-imposable, are relatively similar.

And, I show here just that the actual conclusions are not terribly much different from the

two groups. One can see that it's a fairly substantial difference in the current number

estimated with Alzheimer's disease, with our group estimating from the combination of

factors that we talked about before, that there's almost twice as many people currently

with Alzheimer's disease as Brookmeyer's group does. However, if we look at the

projection, we will see that Ron's projection is that this number of 2.32 million that he

estimated occurred in 1997 will approximately little less than quadruple over the next

50 years to 8.64 million, and the number of 4.5 million that we estimate right now, we

estimate it will a little less than triple to 13.2 million over the next 50 years.

So, to put this very crudely, whatever your snapshot of Alzheimer's disease right

now, perhaps not in our lifetime, but in the lifetimes of our children, this will change

dramatically. And, I think both groups estimate that the number of persons will increase

greatly over the next 50 years or so.

Just to give you a brief breakdown of this, I've given the figures for the whole

United States now. This is work that is also done by the same people I showed on the

first slide, that's currently in the press and gives these projections by state within the

United States. And the message is really just two-fold. One is that the states vary

enormously, both in the number of people in their populations that have Alzheimer's

disease currently, and in what will happen to these numbers over the next 50 years. This figure shows the categories of the number of people with Alzheimer's disease. In general, this is highest in states with currently the oldest populations, so the states in the industrial northeast tend to have pretty substantial number of people with Alzheimer's disease. Those states that people tend to retire to when they become older tend to have substantial number of people with Alzheimer's disease. If we turn this now and look at the growth that's anticipated over the next 50 years, we'll see that the growth in those states that have the highest numbers, with some exceptions like Texas, is going to be relatively smaller, okay, that is, that the Illinois, New York, Pennsylvania are not going to change terribly much. They're in the zero to 20% group. The growth is going to be very large in states like Texas and also in states like Florida, although they're in different categories of growth, to which people retire when they are older, and in those states with the young populations right now, because they're started from the low base. And, as those populations age and there is migration to them, they are going to become older.

Thank you. That concludes the talk. (Comment off microphone) Oh, I'm afraid there's not too much money in estimating the number of people with Alzheimer's disease, and I have none – (laughing).

Neil Buckholtz, Ph.D., National Institute on Aging

Thanks, Denis. There's one more presentation in this group, which is going to come later from John Trojanowski. And, so, the discussion that John Morris will be leading again later on this morning will cover this area.

At this point, we would like to switch over to the PET area and, as I said, we'll have a separate discussion on that afterwards. And, the first speaker here will be Kirk Frey from the University of Michigan, who'll be talking about what does PET add to the clinical diagnosis of AD?

Kirk Frey, M.D., University of Michigan Hospital

Great, thanks for the invitation to be here. I must say I'm actually representing Gary Small's group at UCLA, as well. They were unable to attend because of issues with the religious holiday.

I wanted to talk a little bit more broadly, at least initially, on PET and its application to dementia and Alzheimer's disease. It's unfortunate that John's talk hasn't preceded mine, but we need to keep in mind, at the present, that Alzheimer's disease is a clinical and pathological entity but, still, we do not have a clear definitive understanding of what the pathophysiology that leads to the changes we see is. And, this places it at a great disadvantage in terms of crafting targeted interventions that might alter the risk of Alzheimer's disease or the possibility that we might be able to modify its course. With positron tomography, we have the advantage of being able to study a number of

25

hypothetical mechanisms that could be involved in Alzheimer's disease and/or its evolution, and I'll illustrate a few of them for you.

So, I'm sure that you're all well aware of the amyloid cascade hypothesis in degenerative dementia. A recently introduced method from investigators at the University of Pittsburg now allows the depiction in the living brain of the amyloid burden and, according to a study just reported in the *Annals of Neurology* from Klunk and his colleagues, is potentially able to distinguish the amyloid burden in patients with MCI from the greater burden in patients with established Alzheimer's disease.

I'm sure that you're all well aware, as well, of the cholinergic hypothesis in Alzheimer's disease from a decade or more ago. Investigators from our laboratories have focused on depiction of changes in the basal forebrain cholinergic nuclei and the presynaptic terminals in the cerebral cortex and hippocampus, arising from them with positron tomography. And, here what you see are surface view cartoons from average populations of Alzheimer patients. The {PNP K3}, the second row, depicts the activity of acetylcholinesterase. And, what's colored here is the reduction compared to age-comparable normal populations. So, you can see that there is a global reduction averaging more than 30% in cholinergic enervation of the cerebral cortex depicted in this way. The next row is the binding of iodobenzyl vesamicol to presynaptic cholinergic vesicles, the presynaptic vesicles containing acetylcholine. And, again, you see a broad abnormality extending throughout the cerebral cortex but sparing the cerebellum and

subcortical structures. And depicted on the bottom is the typical pattern of FDG

hypometabolism that was identified in these same individuals.

What about the role of inflammation? You're all aware, I'm sure, of

epidemiological evidence that suggests that anti-inflammatories and their exposure may

reduce the risk of developing Alzheimer's disease. This too can be tested with positron

tomography. And, here, from the group at Hammersmith Hospital is depicted a

substantially increased finding of PK11195 to the so-called peripheral benzodiazepine

site. And, this is put forth as evidence of active inflammation, probably microglial

activation in patients with Alzheimer's disease.

So, this brings us finally to the subject of this morning's discussion, and that is

what about detecting dysfunction and loss of cortical association neurons in neuropil.

And, so, here is a cartoon of acidic amino acid neurons and their projections in the brains,

and you notice those cells, the cortical association neurons that are predominantly

glutamatergic, are known to be selectively depleted in autopsy analyses of Alzheimer's

disease.

In our laboratories, we've looked at the impact of Alzheimer's disease on the

general distribution of synaptic neuropil. And, here what you see is carbon-11 labeled

flumazenil binding to the GABAa receptor, a predominant post-synaptic receptor

distributed throughout the brain. And, what you should see here is that in Alzheimer's

disease, the bottom row, there is really no significant decline in this marker of synaptic

neuropil and, in fact, contrasted against the abnormalities in 18 FDG, one can see that

there is synaptic dysfunction measured by FDG in areas where a synaptic structure, at

least as measured by GABAa receptors, is relatively well maintained. Now, we do not

have selective PET measures yet of presynaptic glutamate nerve terminals, but I would

put forth that FDG is predominantly sensitive to excitatory neurotransmission in the brain

and, thus, numerous investigators who have reported on this distribution of cortical

hypometabolism in clinically diagnosed Alzheimer's disease involving predominantly the

association cortices in the parietal and temporal regions, more so than in the prefrontal

cortex, and as noted by Minoshima and colleagues, when he was at University of

Michigan, a very prominent and early involvement of the posterior cingulate cortex in the

medial aspects of the hemisphere. And, in fact, Minoshima reported in patients with

isolated memory impairment, which is psychometrically slightly distinct from amnestic

MCI, it was possible to show that this reduction in the posterior cingulate was present in

advance of the ability to make a clinical diagnosis of dementia.

Is this really related to what we think is the primary early pathology in

Alzheimer's disease? Well, Minoshima and colleagues went on to study this in terms of

surgically induced temporal lobe dysfunction. So, this is a collection of patients with

left-sided temporal lob epilepsy, who were scanned before and after their anterior

temporal lobectomy. And, depicted in red in the hot metal scheme here are the areas in

brain where metabolic reduction was seen after temporal lobectomy. And, of course, in

the missing temporal lobe, there's no metabolic activity whatsoever, but if you look to

the far right in the left hemisphere on the medial projection, you can see that there is an induced reduction in metabolic activity in the posterior cingulate, simply from surgically resecting the hippocampal formation.

This leads me to two clinical trials that evaluated the possible impact of positron tomography in the evaluation of suspected dementia. The first of these comes predominantly from the investigators at UCLA, together with their international consortium, where subjects undergoing evaluation for dementia were supplemented with FDG-PET. This was a retrospective series, including a clinically followed cohort from UCLA and an autopsy confirmed cohort from the other referring institutions. These investigators classified the FDG-PET scans according to the presence of a neurodegenerative pattern of abnormality and also specifically for the presence of the Alzheimer's disease pattern. These data were then compared with histopathological data that were additionally classified for the presence of any nerve degenerative {dementing} disease and also specifically for Alzheimer's disease.

And, this cartoon sort of depicts the categories that were established, so on your left are what were considered to be non-progressive patterns of metabolism, including normal or global hypometabolism relative to the basal ganglia or specifically focal areas of hypometabolism as may be seen, for instance, in cortical stroke. On the right are the progressive patterns that were identified. The first of these, parietotemporal with or without frontal hypometabolism, the second of these, frontal predominant

hypometabolism, and the third pertaining to Huntington's Chorea bilateral striatal hypometabolism.

The results are depicted here. The sensitivity, specificity and accuracy with regard to the presence of a degenerative dementia were really quite good. With regard to the specific presence of Alzheimer's disease changes at autopsy, they were equivalently good, which argued, strongly I think, that the diagnostic information available in FDG-PET is outstanding. When they went back and looked at the population of patients in whom clinical diagnoses were available from chart review, on your left, there are, before considering the use of PET, the likelihood of progressive dementia or Alzheimer's disease considered by the clinician, "P" for positive, "N" for negative, and "indeterminate" for clinically uncertain. You can see here that there was a considerable population of patients in whom there was diagnostic uncertainty or error who then clinically progressed and developed Alzheimer's disease, depicted in red. If one included the results from FDG-PET, you see the result in the two bars on the right, where now diagnostic confidence in the presence of a progressive dementia was considerably enhanced, and the indeterminant group was eliminated.

Now, a study that was done at our center, together with Norman Foster, Dave Kuhl and Satoshi Minoshima, involved a prospective design. And, here, we enrolled patients when they presented, at their initial visit, for a new cognitive complaint or abnormality. There were two referral clinics; one was a specialty cognitive disorders clinic, and the other was a general neurology clinic at our Veterans Administration

Hospital. We used a standardized clinical evaluation, which included all the routine measures recommended by the Academy of Neurology. We then conducted a two-year follow-up period to establish a clinical diagnosis of progression versus a non-progressive problem. We included PET imaging at the entry session, and we used an automated objective analysis to determine whether the scan had an AD pattern, that is hypometabolism in the association cortical regions, and we required for the –

(End of Tape 1, Side 1)

– a patient as positive for AD, that this pattern be present independently in both cerebral hemispheres. And, this is the flow in the patients, so we enrolled 116. We had complete clinical follow-up and/or autopsy results in 90 of these. And, depicted on the bottom are the performances here regarding progressive versus non-progressive disease, very similar to what we've seen at UCLA. And, if we look at the subgroups that satisfied clinical criteria for "normal," that is, no neurologic illness versus those who made ADRDA criteria for probable AD, we see that there's an even greater ability to discriminate these.

We did have some false positive scans, that is, those that had an AD-like pattern, but who did not have progressive cognitive decline. All of these subjects had abnormal neurological or neuropsychometric examinations, so they were not normal subjects. The clinical diagnoses here were multiple sclerosis, and the results in the scan was because of peritrigonal large confluent plaques that effectively denervated the {association} cortices,

31

or subjects, largely from our VA hospital population, who had a history and continuing problems with ethanol abuse.

False negative scans, that is, scans that did not depict the AD pattern but subjects who did have progressive cognitive decline nonetheless, well, none of these actually had a normal FDG pattern. They all demonstrated unilateral, i.e., one hemisphere but not both, abnormalities. And, this was not simply due to missing earlier mild cases, in that the CDR ratings for these patients varied from one to two. We included in this cohort a number of subjects with CDR 0.5, and none of them were the false negatives I'm reporting.

So, in conclusion, these two studies indicate that FDG-PET is sensitive and specific for the identification of progressive cognitive decline, both retrospectively and in prospective, unselected populations presenting for the evaluation of a cognitive complaint. Sensitivity is very high, and specificity is, in at least our study, virtually 100% for distinguishing probable or definite Alzheimer's disease. And, we think from these data that, combined with clinical and neuropsychometric data, FDG-PET offers a highly accurate clinical diagnosis in suspected dementia.

Now, when considering the possible impact of FDG in clinical practice, we think that there are at least three aspects that need to be considered, and some of these are absent from a number of the analyses that have been published. First and, of course, not to be downplayed, the cost of a diagnostic evaluation, but very important aspects that are

not routinely included in these cost analyses include the value of accurate clinical

diagnosis. It's critical to the clinician, interfacing with the patient and their family, to be

able to provide an accurate diagnosis so that a prognosis can be given, so that patient and

family counseling regarding the expectation of evolution of the problem can be made,

and so that the patients and their families can engage in future planning. The validity of

these acts is entirely dependent on how secure one is in the diagnosis. And, I would draw

an analogy here to 25 years ago in clinical neurology, the diagnosis of multiple sclerosis,

demyelinating disease. We were often confronted with patients in whom the diagnosis of

MS at their first presentation was considered by the clinician as the most highly likely

diagnostic entity. But, because the diagnostic certainty required that the patient

experience multiple attacks over time, patients often had to seek multiple evaluations

from multiple neurologists, and experience a waiting period of a year or two, before

a diagnosis was actually given. The advent of MRI in the depiction of silent but

historically previous MS episodes greatly revolutionized our ability to make this

diagnosis, often on the first clinic visit. I think that FDG-PET has the same opportunity

with regard to Alzheimer's disease.

Finally, an issue to consider is the impact on treatment. And, here, not only the

risk/benefit ratio but also the cost/benefit ratio needs to be considered. There are

presently two classes of medication FDA-approved for the treatment of Alzheimer's

disease. Each one of these may cost a patient $1,000 to $2,000 a year in out-of-pocket

medication expense. So, to recommend that we treat all of our patients who present with

an amnestic complaint is quite a financial burden that we have displaced to the patient. There's also the consideration of risk/benefit ratio, as illustrated by the immunologic trial that was reported a year or so ago by Elan Pharmaceuticals, where immunologic therapy, aimed at arresting the progression of Alzheimer's disease, actually resulted in permanent neurological deficits in I think it was eight patients. We would certainly not want to expose subjects without disease to the risk of such a detriment.

Now, Silverman and his colleagues have actually conducted a comparison of FDG in the diagnosis and management of Alzheimer's disease, and their model included expenses of treatment in the assessment. They're very complicated to depict but very easy to explain, clinical diagnosis (inaudible) where they've included PET only for subjects in whom diagnostic uncertainty remains after the routine application of usual clinical diagnostic procedures. What they concluded was that the rate of incorrect diagnosis was cut in half by the introduction of PET, compared to the routine clinical conventional algorithm, and that this had a favorable financial impact, saving thousands of dollars in evaluation and in management expenses. According to their analysis, the cost neutral entry level for PET was around $2,700, which is considerably in excess of the expense at most centers today.

So, the conclusions I'd like to leave with are that certainly PET and SPECT neurochemical imaging in dementia research has promise in areas of assessing pathophysiological change in the potential to identify mechanisms of therapeutic

intervention and possibly also to track disease progression in neuroprotection. And, we'll

hear about some of these aspects, I think, from Dr. Reiman.

Roles of FDG-PET in clinical assessment include as a diagnostic aid in the

evaluation of presence of neurodegenerative dementia when there is question. In

addition, there may be great value in differentiating among subtypes of

neurodegenerative dementia, as we've heard already and as will be illustrated further by

Dr. Foster. Thank you.

So, I have no conflicts, other than I accept, on behalf of the university, payment

from CMS for my patients – (laughing). I believe Dr. {Phelps}, who I listed as a

coauthor, has a financial interest in CPI PET systems a manufacturer of PET tomograph.

Neil Buckholtz, Ph.D., National Institute on Aging

Thank you. Our next speaker is Eric Reiman from the University of Arizona.

He'll be talking about PET in early detection and tracking of Alzheimer's disease.

Eric Reiman, M.D., Good Samaritan Regional Medical Center, Phoenix, Arizona

Thank you, Neil. While I'm getting set up, I am involved in PET research and

Alzheimer's disease research. I am not formally involved in the clinical evaluation of

patients with dementia with PET, and I have no conflicts with regard to today's

presentation.

What I'd like to talk about today is the use of FDG-PET in the early detection and tracking of Alzheimer's disease and how it may be relevant to the differential diagnosis of Alzheimer's disease and the prediction of a clinical progression in cases with suspected dementia, in cases with MCI, and in cognitively normal individuals at risk for the disorder. That's what I'd like to talk about today. Is my 20 minutes up yet?

While we're waiting, I'll begin without some images. FDG-PET, as you've heard, reveals some characteristic and progressive reductions in the cerebral metabolic rate for glucose. This includes reductions in the posterior cingulate, parietal and temporal cortex, later in the course of the illness, reductions in prefrontal cortex and in whole brain metabolism. These changes in cross-sectional studies are correlated with dementia's severity and, in longitudinal studies, appear to be progressive. As you've heard and as I'll discuss a little bit more, these findings tend to predict clinical decline in the histopathological diagnosis of Alzheimer's disease, in patients with dementia. In groups of individuals, they predict rates of conversion to probable AD in patients with MCI. And, we and others have begun to detect and track progressive reductions at each of the same brain regions in cognitively normal carriers of a common susceptibility to the ApoE4 allele.

Our interest in the study of those cognitively normal individuals and, indeed, in our longitudinal studies of patients with dementia has been, in part, to help provide a surrogate marker that could assist in the discovery of drugs to treat and prevent

Alzheimer's disease, providing smaller subject samples for proof of concept studies and to help assess the disease-modifying effects of candidates.

You know what? We tried that, and the font wouldn't convert. That was the problem in there. Thank you, appreciate it. Well, this study utilizes data kindly provided by Satoshi Minoshima and his former colleagues at the University of Michigan, and it's a surface projection statistical map showing significant reductions in 37 patients with probable Alzheimer's disease versus 22 normal controls, and you can see the reason {for the preferentially} affected in patients with probable AD. It's been suggested by Dr. Minoshima and others that reductions in the posterior cingulate might be particularly sensitive to detecting the changes early in the course of the disorder.

In cross-sectional studies conducted by that group, one can see correlations with disease severity. Looking at the column in your far right, you can see, looking at these maps, a number of standard deviations below the mean of normal controls, progressive reductions in these brain regions in cases of moderate versus severe Alzheimer's dementia. And, longitudinal studies involving our group, we've been able to track these changes in these regions over a one-year time and estimate the number of patients needed to test the disease-modifying effects of putative treatments, suggesting that PET could be used, not unlike MRI measurements of hippocampal volume in whole brain atrophy, in smaller subject samples and proof of concept studies to assist in the assessment of putative disease-modifying treatments.

When it comes to the role of FDG-PET in the differential diagnosis of patients

with respect to dementia and in the prediction of clinical decline, I agree with Dr. Frey

that the most informative study comes from Dan Silverman and his colleagues. And, in

this study, they looked at 146 patients who had clinical follow-up at UCLA and an

additional 138 patients from eight academic centers who were followed subsequently

with autopsy an average of 3.2 years after the onset of PET. This study is particularly

informative for patients early in the course of the illness, as 70% of them had evidence of

mild or questionable AD, and the majority of these individuals had Mini Mental State

Exam scores in the range of 26 to 30.

In addition, these individuals were not uniform. They didn't all go on to have a

progressive course; they didn't all have Alzheimer's disease; and it was helpful to

determine the role that PET would play in help predicting progressive versus

non-progressive clinical decline and predicting the histopathological presence or absence

of Alzheimer's disease. In their studies, they reported 93% sensitivity in correctly

predicting progressive clinical course, 76% specificity in correctly predicting a

non-progressive clinical course. They found a similar sensitivity and specificity in the

prediction of the presence or absence of histopathological Alzheimer's disease. And,

these numbers were similar if the data were restricted to those with that earlier courses

with a mild or questionable AD. To my surprise, quite frankly, they found a 94%

concordance rate between two PET readers using visual inspection and a very simple

couple of paragraph description of the criteria used to evaluate these scans. It'd be nice

to be able to report the concordance among readers outside of the study.

The study had a number of limitations, although I still believe to be a very

informative study, suggesting the need for additional information, which one might

consider getting in post-approval field trials. One needs to directly compare the PET

findings with the clinical findings of dementia specialists and/or primary care physicians.

That was not done in this retrospective study but was partially addressed in a subsequent

study in which they looked at those individuals retrospectively in terms of the prediction

of the presence or absence of AD and the prediction of clinical decline. And, they found

that PET was particularly helpful in those patients who were clinically determined to

have non-progressive clinical decline or indeterminate clinical decline in predicting

depression. It also helped in a few cases that were thought to be progressive in predicting

non-progressive decline, but it was most helpful in quite a number of individuals in

whom they were not thought to be progressive, just clinically.

While there's indirect reason to suggest that the improved diagnostic accuracy

with PET will be of benefit to patients and their families, that work needs to be done

directly to show the impact of this modality on the emotional well-being of patients and

their families, how much just reducing the uncertainty of this catastrophic diagnosis help

with the emotional distress of patients and their families. How much does it help them

make the proactive judgments used to care for their loved ones, take the car away, and

plan for their future? What effects does it have on access to treatment, the appropriate or

inappropriate use of medical treatments, the postponement of nursing home placement? While there's reason to believe, both based on anecdotal experience and indirect studies of the benefits of providing intervention and support on delayed nursing home placement, one would like to see the direct impact of the diagnosis with PET on these outcomes.

One would like to see the findings extended to other patient samples. The samples in this study, these were academic centers. They were slightly younger patients overall than one might see in community settings. One would like to see results extended to those community settings. One would like to see the reliability of clinical interpretations based on visual inspection of the data extended to other PET readers, and would like to see direct comparisons of visual interpretations to automated image analysis techniques, which may turn out to have even more power, something that Dr. Foster will talk about in the differential diagnosis between Alzheimer's disease and frontotemporal dementia.

To illustrate the use of statistical image analysis techniques, those purple areas were the regions of the brain found to be preferentially affected in patients with Alzheimer's disease. We can create Z-score maps to show the pattern of reductions in individual patients, superimpose them on this map, and show the regions in blue that conform to Alzheimer's disease, providing – we found this helpful in those individuals who were overwhelmed with this catastrophic diagnosis and remaining uncertainty, were not taking the proactive steps involved in their care, and could use this information educationally as well as diagnostically.

One can use other techniques, as well. Here's a patient known to carry an Alzheimer's causing gene, who was clinically affected with dementia. One can see in this horizontal slice these reductions in the temporal-parietal cortex, but one can with even greater confidence the Z-score map using statistical parametric mapping, showing the characteristic pattern found with patients with Alzheimer's disease, at 14 standard deviations below the mean in that individual.

In terms of the relevance of FDG-PET to the prediction of clinical progression in patients with MCI, there are two small studies that were done, suggesting the promise of FDG-PET but really raising the need for additional studies before any definitive conclusions can be made. There was a small study of patients with non-amnestic MCI. Twenty patients were followed on average just over three years later, but a wide range in the follow-up period. And, they looked specifically at left and right temporal-parietal regions of interest. They found that a left temporal-parietal region correctly classified about 75% of patients, but they didn't look at other regions, small numbers and not enough to warrant clinical use at this time.

In a complementary study, investigators looked at 17 patients with an average of 18-month follow-up, again not a consistent follow-up period. And, they use statistical parametric mapping to survey the entire brain, predicting that they would see changes in temporal-parietal and posterior cingulate regions. And, when they looked at their temporal-parietal and posterior cingulate regions of interest and chose the voxel that best distinguished the group, they saw a good distinction between those who subsequently

converted – the probable age it'd be from those who didn't convert, in these regions, but

they found it primarily in the right temporal-parietal region. They didn't find much

sensitivity in the left temporal-parietal region. They also found it in the posterior

cingulate region. And we need much larger sample sizes and prospective criteria in terms

of regional comparisons to determine the accuracy in predicting probable Alzheimer's

disease before this would be clinically warranted.

Our group and others have been interested in detecting and tracking these brain

changes in individuals at genetic risk for Alzheimer's disease. We've been capitalizing

on studies of cognitively normal persons with two copies, one copy and no copies of the

ApoE4 allele, a common Alzheimer's susceptibility gene, or individually matched for

their age, gender and educational level. In our original studies, we found, as shown

in blue, that a cognitively normal ApoE4 homozygous 50 to 65 years of age had

significantly reduced glucose metabolism in each of the same brain regions as patients

with probable AD. We've extended this work to ApoE4 heterozygous, as well. In our

original study, the person who had the largest reductions in each of these regions, his

wife called back six months later to report cognitive concerns. He was in a nursing home

two months later. That said, there's significant overlap between groups, and we still need

to do the longitudinal studies to determine the extent to which these changes predict

subsequent cognitive decline, onset of MCI, and probable AD.

In a complementary study of individuals about 10 years old, those who present to

a memory clinic with memory concerns and had slightly lower Mini Mental State Exam

scores, Gary Small and his colleagues found that some of these changes, particularly in parietal cortex, predicted two-year declines in memory. But, we do not believe that PET is indicated to predict a cognitively normal person's risk for developing Alzheimer's disease. They don't tell us with sufficient certainty whether somebody will develop Alzheimer's dementia, when they might develop it, what we can do about it at the present time, what to tell their employer or their third-party payer. I think of it more as a research tool.

We're now following 160 ApoE4 homozygous, heterozygous {to} non-carriers longitudinally to determine how baseline measurements and two-year declines predict a subsequent onset of MCI and Alzheimer's disease. And, we do believe that these reductions will have some significant predictive role, though not necessarily clinically significant role, in predicting subsequent declines. Here, you see the correlations between significantly lower rates of glucose metabolism in each of the brain regions affected by Alzheimer's disease. A number of copies of the ApoE4 allele are independent risk factor for the risks of Alzheimer's disease, and here you can see the numbers as well as the overlap among groups.

We've recently extended these studies to young adult carriers of the ApoE4 allele. These are ApoE4 heterozygotes several decades before the onset of symptoms, a finding that raises a number of questions. Are these changes very early age-related changes? Are they neurodevelopmental changes? How are these changes related to the susceptibility to the histopathological features that develop a little bit later in patients

with Alzheimer's disease? To what extent can they be found in non-carriers of the

ApoE4 allele who are at risk for the disorder? But, what I wanted to be able to show you

was the significant overlap between groups. We don't how theses individual changes

predict subsequent decline, and they're not indicated clinically to predict a person's risk

for developing Alzheimer's disease. Indeed, none of the individuals in our studies get

any information about their individual PET finding.

To add further a note of caution to the prediction of risk for the disorder is a study

that we did in conjunction with my colleague, Rick Caselli from the Mayo Clinic in

Scottsdale, of a neurodegenerative disorder that sometimes has the histopathological

features of Alzheimer's disease, cortocobasal degeneration. We looked at identical twins

who were clinically discordant for this disorder. And, as you can see in the top row, the

affected twin had significant reductions in whole brain metabolism, particularly in the

temporal-parietal regions. The unaffected twin also had, in comparison to normal, a

somewhat lower whole brain metabolism. And, as you can see in the section in the

middle, a reduction in temporal-parietal glucose metabolism, actually in the left, he also

lower language scores than his genius level, non-verbal scores. And, that was eight years

ago, and this person eight years later is still fine. He hasn't progressed. It wasn't

clinically predictive. This person didn't get the results of his images, either.

While amyloid imaging has promise in the early detection and diagnosis of

Alzheimer's disease, more work needs to be done to help refine and test the quantitation

procedures and apply them in the proper way to determine what their role may be.

While, ultimately, there may be some value in using complementary modalities, imaging and non-imaging diagnostic tests in predicting those individuals who are greatest risk for developing MCI and AD.

In conclusion, I'd like to suggest that PET, FDG-PET may provide helpful information in the differential diagnosis of AD and the prediction of progressive clinical decline in patients with suspected dementia. But, while FDG-PET is promising, it is not yet clinically indicated in differential diagnosis of AD or the prediction of progressive cognitive decline in patients with MCI. Longitudinal studies and larger samples are needed to help clarify its clinical role in these patients, whether used alone or in conjunction with other diagnostic measures. And, it's not clinically indicated to predict the cognitively normal person's risk of developing MCI or AD.

On a personal note, I would support the consensus recommendation made by the Alzheimer's Association and suggest that reimbursement for the appropriate use of FDG-PET in patients with suspected dementia in whom there are uncertainties following a comprehensive clinical assessment. I would suggest that additional steps needed to optimize the appropriate use of the study, including the use of a referring physician friendly order form to ensure that a comprehensive clinical exam has been performed and to minimize the inappropriate use of PET as a substitute for other indicated management plans which might include referral to a specialist. I would also suggest the certification of PET readers, who understand the criteria used to interpret PET scans and who demonstrate adequate reliability in reading a set of training scans. I've been a little bit

concerned, just based on anecdotal experience, of variable criteria used at less experienced sites.

I'd also suggest the need for field trials after approval to further support the added value of FDG-PET, both the primary care physicians and dementia specialists, to monitor the reliability of PET readings and to identify image analysis techniques with even greater diagnostic accuracy. And, I think it's important to consider now the extent to which the proposed neuroimaging initiative will be sufficient to help determine whether or not to reimburse the appropriate use of FDG-PET, MRI, the combination of these techniques in patients with MCI and, if not, to consider now what additional studies need to be done.

I'd like to thank my colleagues for the studies performed by our group, and I'd like to thank you for your patience.

Neil Buckholtz, Ph.D., National Institute on Aging

Thank you, Eric, and introduce Norm Foster from the University of Michigan, who'll be talking about PET in differential diagnosis of select cases of dementia.

Norman Foster, M.D., Department of Neurology, University of Michigan

Thank you. Susan Molchan asked me to talk about use of PET in select cases. And in my own personal experience, I have many cases that I can tell you about in which management has changed. Sometimes, inappropriate medications were discontinued and

appropriate medications begun, in other cases, where management was changed, as well as patient and family approaches to their disease was changed. And, in some ways, I think that these individual cases are most convincing. But, today, what I want to do is tell you about a study that I was directing, which tried to answer a practical clinical question. In fact, these individuals' cases or anecdotes that I have told you about also had the overall theme that the use of PET imaging is helpful in answering specific clinical questions. And, the clinical question that we were interested in is one that physicians ask themselves thousands of times when they are sitting in front of a patient: "Does this person have Alzheimer's disease, or do they have frontotemporal dementia?" This isn't an easy question to answer, as we've already heard from Ron. Most FTD patients meet criteria for Alzheimer's disease.

In addition, the clinical judgment is based upon very subjective decision about whether memory is more prominent or behavioral and language problems are more prominent. And, in fact, FDG-PET helps with this same kind of process that clinicians are going through, which is trying to understand where pathology is primarily localized in the cerebral cortex. So, we thought that FDG-PET might help in this clinical distinction. In order to do this, we were able to receive funding from the National Alzheimer's Coordinating Center, and we engaged three different Alzheimer centers, including the Coordinating Center, to evaluate in a retrospective study whether this could be helpful. What I'll be reporting about today is 45 autopsy confirmed patients who had either Alzheimer's disease or frontotemporal dementia; 31 of them had Alzheimer's

disease, and 14 had frontotemporal dementia and only that pathology at autopsy. And, as you can see, patients were scanned throughout the course of their dementia at various times.

We evaluated four different diagnostic approaches to distinguishing Alzheimer's disease from frontotemporal dementia. We had an expert neurologist who specializes in dementia work who reviewed the entire medical records of the patient, the history and examinations, excluding any imaging results. She was blinded to the diagnosis and pathologic findings and developed a one- to two-page scenario summarizing the patient's clinical course. And that was used for one method. The second was a checklist based upon this scenario that has been published by Barber. And, then, we looked at two different methods of looking at PET images, the standard method of transaxial images and also stereotactic surface projection maps that you've been seeing today, also. These different items were sent separately to these six raters, and they were asked to determine whether the patient had FTD, or decide whether the patient had frontotemporal dementia or Alzheimer's disease, and indicate their degree of diagnostic certainty. They knew already that these cases had one or the other, but they did not know the proportion of each.

Now, just to give you an idea about what these images look like, remember they were sent either transaxial images at what point or at another point, stereotactic surface projection maps. This is a normal individual and, in these SST images, we showed either relative glucose metabolism compared to pons here, so a metabolic map. And,

secondarily, we showed a statistical map where we compared that specific individual with 33 normal control subjects and, so, we were able to display a Z-score.

We used a simple and practical clinical visual interpretation of the scan, simply saying that those who had Alzheimer's disease had predominant hypometabolism in association, temporal-parietal cortex, and posterior cingulate cortex, while those categorized, and we asked them to categorize the scans as frontotemporal dementia had predominant frontal anterior temple and anterior cingulate hypometabolism. That was the simple rule. The raters were six dementia specialists from three Alzheimer's centers funded by the NIA. There were two at each of the centers, and these individuals, even though they all were experienced dementia specialists, had variable experience with PET, some of them recognized in the field, others complete novices. So, we carried out a training session over the telephone, in which we reviewed image methodology and anatomy and, then, reviewed and discussed a set of FDG-PET images that were not otherwise used in the study from 10 clinically diagnosed patients with Alzheimer's disease, 10 with FTD, and five normal elderly control subjects. And, these cases were selected to show the entire range of abnormality that we expected to see during the course of the study.

So, that was the study, and the first thing we wanted to do was to determine how reliable diagnoses were. This is a very clinically important issue, because patients with dementia often go between one physician and another and get different answers what the cause of the dementia is. And, there can be some disagreements, and we could see this in

our own study. So, first, we're looking for unanimity among all of the raters in the

diagnosis or a supermajority. And, as you can see, there was significant frequency of

people who did not agree on the diagnosis. The clinical methods performed less well

than the PET methods did. We also looked at inter-rater kappa scores; the average kappa

value was actually significantly worse for a checklist diagnosis alone but was higher with

both PET imaging methods than with the scenario, the SSP method being somewhat

better. Next, we looked at diagnostic accuracy, again unanimity, supermajority, overall

accuracy. With each of these measures, as you can see once again, the clinical methods

were less accurate than were the PET methods and had less inter-rater variability in

accuracy, also.

So, now, I want to talk about the usual ways that we assess the value of a test, with

sensitivity, specificity and predictive value. First, looking at Alzheimer's disease. And

you realize there was a decision to be made between Alzheimer's disease and

frontotemporal dementia. So, these two are kind of obverse of each other. As Ron has

already told us, clinical sensitivity for the diagnosis of Alzheimer's disease using clinical

information is quite good, but specificity is lower. These are values within those that

Ron reported. But, once again, as you can see, PET methods were in general superior to

clinical methods for the sensitivity and specificity of Alzheimer's disease. Just again, as

Ron had mentioned, the diagnosis of frontotemporal dementia is quite specific, using

clinical methods with lesser sensitivity. And, again, checklist actually helped the

specificity but hurt the sensitivity. But, the specificity and sensitivity was improved by or was better in the imaging methods than the clinical methods in the study.

The predictive value for Alzheimer's disease, positive predictive value was somewhat higher with the imaging methods. Negative predictive value was better, particularly with SSP imaging, and many people now use the positive likelihood ratio to determine the added value of the PET, and because of the high pretest probability of Alzheimer's disease, the positive likelihood ratios are not very high with Alzheimer's disease. For frontotemporal dementia, we see the same pattern once again with SSP being the superior method. But, in this case, because of the low prevalence of frontotemporal dementia, positive likelihood ratio for frontotemporal dementia is really extremely high, anything greater than 10 generally felt to be a compelling value of a diagnostic test. So, either transaxial or SSP image methods meet that standard.

So, those are traditional values or ways to assess a test, but I want to also present interesting information from this study that talks about another way, which we've heard a little bit about before, which involves how confident physicians are when they are speaking to patients about the diagnosis. My own view is that this may make a major difference in how likely physicians are to give a patient and family a diagnosis, discuss a prognosis, or the likelihood of their implementing an appropriate treatment. And, so, what I am showing here are incorrect diagnoses in blue and correct diagnoses in red, demonstrating on the intensity of the color the confidence. So, we asked the raters to determine the confidence. Were they very confident, somewhat confident, or uncertain?

51

We forced them to make a diagnosis of AD or FTD, but they might have felt uncertain about that, even if they were correct. And, so, I'm comparing here the certainty and correctness of scenarios versus stereotactic surface projection PET with FTD and AD patients. And, although using the typical ways to assess the value of a test, we saw that FTD was much more compelling. You can also see that in patients with Alzheimer's disease that there was a sizeable improvement in the diagnostic certainty of the ratings of diagnosis based upon the PET scans as opposed to clinical information. So, we decided from this, we can conclude from the study that FDG-PET can improve the accurate diagnosis, particularly of frontotemporal dementia, and that this is possible after relatively brief training, using a practical, clinical method of assessing what the cause of the dementia is in interpretation of the scan. This accuracy was better with both FDG-PET methods than with clinical assessment alone or a symptom checklist. It appears that SSP images may have some benefits over the traditional transaxial image method.

We went on to assess, using these same raters and the same cases, the serial assessment. What is the effect of adding FDG-PET to a clinical assessment? So, we had a web-based rating system in which we could show the raters the clinical scenario, lock in their diagnosis and diagnostic certainty, and only after that had been done, show them the SSP PET image and ask them again to make a diagnosis with their degree of confidence. When FDG-PET was added to the clinical scenario, overall diagnostic accuracy increased from 80 to 89%. FDG-PET changed the diagnosis in 16% of the

cases, almost all of which, 81% of the time, corrected a misdiagnosis. All raters similarly

benefited from this FDG-PET, with 13 to 20% of the diagnoses changed, depending upon

the individual rater, and FDG-PET increased diagnostic confidence just as before. So,

here is the same kind of table showing FTD and AD cases, and this is before and after

PET. And, as you can see, there was a substantial increase in both accuracy and

particularly in confidence of diagnosis by adding PET.

Now, there's some strengths of this study, because we used pathologically

confirmed cases. We used expert clinicians who were likely to make clinical diagnosis

more accurate than in community physicians, and I point out that we don't really know

how often community physicians are accurate in making a diagnosis of Alzheimer's

disease or frontotemporal dementia, and these results would be biased against showing

a benefit of PET. It also showed that there was good inter-rater reliability and have

information about what image presentation method might work best. However, it does

have limitations, because the cases studied were highly selected, and scans were not

obtained during an initial assessment. You may have recognized that some of the

patients in this study were actually scanned before they had clinical dementia,

particularly those with Alzheimer's disease, based on a Mini Mental State score, for

example, but some of them also were severely impaired. Likewise, the clinical scenarios,

even though we tried very carefully to give a full picture of the disease, may not be as

good as in-person assessments.

So, as a result, the National Institutes on Aging has funded a study, which I direct, a three-year prospective pilot trial at nine Alzheimer's centers, in which we will be evaluating 54 subjects with suspected and uncertain frontotemporal dementia enrolled at the time of their initial evaluation. We'll be evaluating diagnosis before and after PET. They'll be followed for 18 months by a physician who is unaware of the FDG-PET results, and this information will be used by a consensus panel to determine a gold standard diagnosis for comparison. It's actually a very interesting study, because the question in part has to do with how do you identify frontotemporal dementia suspicion? We're not expecting everybody in this study to have frontotemporal dementia. Indeed, we want and expect that there will be patients with Alzheimer's disease and psychiatric disease included, but we are asking the Alzheimer's centers to act like community physicians and do no more than AAN guideline evaluation. And, I point out that the AAN guideline does not recommend neuropsychometric testing as part of their assessment, and that's done routinely by most Alzheimer's disease research centers. Genetic testing also is not permitted in this group, and people who have reason already to believe that they have frontotemporal dementia are not included in this study.

So we're going to be looking not only at physician diagnosis before and after PET, but also the treating physician's management plan before and after PET and caregiver satisfaction about the diagnostic process before and after FDG-PET. And, so, I'm hoping that this new study will provide additional data in support of the preliminary evidence

that we have that FDG-PET can be particularly helpful in recognizing and distinguishing patients with Alzheimer's disease and frontotemporal dementia. Thank you.

<u>Neil Buckholtz, Ph.D., National Institute on Aging</u>

Okay, we're going to take a break at this point. Hopefully, John Trojanowski will be coming in by train from Philadelphia, and he will be the presenter after the break, at which time we'll have the discussion on the first three presentations. Thank you.

(Break)

<u>John Trojanowski, M.D., Ph.D., University of Pennsylvania School of Medicine</u>

Well, I do thank you, both you and Susan, for inviting me. This is obviously a very important meeting, and I don't know if I can do justice to the topic you gave to me, the histopathological distinctions among dementias, pathology and clinical course, in a mere 20 minutes, but I tried to put together a PowerPoint that would give you a flavor for where we are right now. And, I will state I have no conflicts of interest concerning the material that I'm presenting here, except I'm passionate about the science – (laughing). And, forgive me for coming a little bit late and departing early. I don't know if anything that I'm going to show has been shown already. I have to get back and work on a grant in Philadelphia.

So, I'll just say that Alzheimer's is heterogeneous at the genetic level, which all of you I think appreciate. We know that there are several different genes already that are

responsible, when mutated, for causing early onset Alzheimer's disease, and that's about 10% of cases. And, then, there are the other risk factor genes, ApoE being the poster child for these other genes that are still in the process of being defined. I would point out that Christine Von Broeckhoven has a paper with Dermaut et al. as first author coming out in *Annals* shortly, showing that a pf1 mutation is the {B1} (inaudible) {3z} mutation that causes FTD. It causes Pick's Disease, no plaques and tangles. So, we're still in the process I think of unraveling the mechanisms of neurodegeneration induced by various mutations, including those that largely are responsible for classic plaque and tangle Alzheimer's disease. And, so, plaque and tangle disease with dementia is Alzheimer's disease; those criteria are very well defined now. We're still struggling with the pathology of mild cognitive impairment, but I think you know we have consensus in many domains for the criteria that states that if the patient's demented and there are abundant plaques and tangles, that we're dealing with Alzheimer's disease, and the tangles are the intracellular amyloids, the paired helical filaments every bit as much an amyloid as the extra {exophiala} deposits of Abeta, and you will see in my last slide that that is an overarching theme of all of these, of many late-life neurodegenerative disorders. They are brain amyloidoses.

It's well known, and this is a slide from Steve Arnold at Penn done many years ago with his collaboratives in Iowa, showing the distribution of neurofibrillary tangles in neuritic plaques. And, this is the sort of typical distribution; temple regions are more involved than primary motor cortex, or occipital lobe or what have you, but the frontal

lobes are affected and the caudal levels of the neural access, brainstem, and spinal cord are less affected.

Having said that, there still are cases that are diagnosed as frontotemporal dementia because of their clinical features and imaging that turn out to be Alzheimer's disease. And, many FTDs are misdiagnosed as Alzheimer's disease, and the pathology turn out to be a tauopathy of one or kind of another or dementia lacking distinctive histopathology. So, we're clearly in need of better diagnostics for this group of disorders.

I won't dwell a lot on the cell biology of plaque and tangles, but it's important to know that all of the proteins that eventually become amyloids and cause disease start off their life doing supposedly beneficial things, that they are evolutionarily favorable, rather than put there to shut down the brain or a species or what have you. So, these are the tau proteins, very well understood from the cell biology point of view being microtubial binding proteins without which microtubials disassemble and (inaudible) transport fails. And, they are normally most abundant in axons, but in the disease state, tau will convert from a soluble group of proteins to insoluble filamentous deposits that aggregate in cell bodies to form neurofibrillary tangles and, also, these neuritic profiles, neuropil threads that, in fact, are 95% of the tau burden. So, if you do morphometric studies and measure where the bad tau is, it's certainly in tangles, but that's only 5% of the burden; 95% is in the neuritic processes, and we feel that that contributes to impaired (inaudible) transport, synaptic failure and so forth. And, here's a picture of some of those paired helical

filaments that form the neurofibrillary tangles. These are isolated from an Alzheimer brain.

And, so, one can construct a schematic like this, which has been out there for a while, whereby tau is pulled off the magenta bars off the yellow microtubials and hyperphosphorylated fibrilized and then deposits either as tangles or neuritic accumulations, leading to the death of these neurons shown to (inaudible) cells here, releasing tau to be accumulating in cerebrospinal fluid which has become one informative and helpful biomarker for Alzheimer's disease. So that the whole tau protein forms filaments in Alzheimer's disease, whereas the Abeta, the second diagnostic hallmark, the peptide that forms the second diagnostic hallmark, the plaque, is the cleavage product of a larger protein.

So, there are several additional steps involved in amyloidoses here. This is the good pathway that cleaves Abeta in the middle, preventing Abeta generation and fibrilization, but through Abeta and gamma secretase cleavages, Abeta is produced. It's produced in everyone in this room normally, but those unfortunate people who are destined to get Alzheimer's disease will accumulate the soluble protein and insoluble fibrils that are recognized morphologically as these amyloid plaques. And, by electron microscopy, they too are filamentous. This is a very low power {EM buse}. They don't see the filaments very well, even in front of them, I suspect. But, these are all amyloid fribiles, much like the other amyloids in the brain.

And, here's a schematic that one can think of as explaining how bad stuff happens and targets for therapeutic intervention. Abeta, the red balls, have released themselves and normally cleared, as shown down below here. But, if aggregation occurs, perhaps through interaction with pathological co-factors or pathological chaperones, the red balls interact with these pathological chaperones to become these yellow sticks, the amyloid fibrils depositing and eliciting a reactive glial and microglial response which can contribute to toxicity, so the name of the game here is to block Abeta production, increase clearance, bust plaques, shut down oxidated damage and so forth.

While there are no other Abeta depositing diseases other than Alzheimer's disease, there's a host of diseases in addition to Alzheimer's disease, in which tangles occur. This is not a non-specific consequence of injury. It's specific to a very well-defined group of diseases, some of which, and these are the ones in yellow, are generically referred to as "tauopathies." I guess I would emphasize, as I alluded to briefly, that a number of these can be misdiagnosed in life as Alzheimer's disease. And, so, we're still struggling with better ways to make the diagnosis, and this is where imaging can help and, of course, more informative biomarkers.

Tau and Abeta I've just discussed with you, and I want to just refer also to another major player in Alzheimer's disease, alpha-trinucleon, associated with Parkinson's disease, obviously, and a group of diseases called alpha-synucleinopathies. But, the deposits of alpha-synuclein in Lewy bodies occur very commonly in familial and sporadic Alzheimer's disease, as well as Down Syndrome. Over 50% of individuals with

these diseases may have Lewy bodies. How they add to or exacerbate the plaque and tangle pathology is not yet clear, but this is a third type of amyloid that accumulates in Alzheimer's disease.

And, then, there are a large group of dementing disorders, many clinically similar to Alzheimer's disease, that have only tau and alpha-synuclein deposits and no Abeta. And, these down here are examples. We have active collaborations with Bruce Miller and Mike Weiner and others at UCSF, as well as with Murray Grossman at Penn, looking at these diseases, and it's still very much an exciting work in progress, sorting out what the diagnostic criteria will be that enable us to talk with more certainty about the classification of these diseases. But, a lot of progress has been made and is being made.

And, so, I'll just say that the third player to think about in Alzheimer's disease and other dementing conditions, Lewy body dementia, for example, which again can be clinically very difficult to distinguish from Alzheimer's disease, are those that result from alpha-synuclein deposits. Here's two of the mutations that were described several years ago. There's another E46 (inaudible) mutation, and we know that there's duplication of the alpha-synuclein gene that's not mutant, that over –

(End of Tape 1, Side 2)

– syndrome. Of course, there, that's the whole chromosome of trisomy 21, whereas in these familial dementia with Lewy body kindreds that have a duplication of the alpha-synuclein gene. It's the region around that gene, not just alpha-synuclein

60

but some flanking sequences, the net effect of which is to increase production of alpha-synuclein. And, the pathologies that we know of that are formed by these predominantly occur in neurons. The most well known, of course, is the Lewy body in the substantia nigra. This is the light microscopic view and an EM view.

But, glial cells constitutively express low levels of alpha-synuclein, too. You can detect them in culture. It's hard to detect them in brain sections. But, in diseases like multiple system atrophy, which so far does not seem to have a hereditary basis, the accumulations of alpha-synuclein and oligodendroglial cells and, intriguingly, this is not only associated with oligodendroglial dysfunction and degeneration, but also neuronal and axonal degeneration, a very intriguing mechanism yet to be elucidated to explain them. Again, as with tau pathology, there's prominent neuritic pathology. We've looked at (inaudible) family members in collaboration with Larry {Goldbe}, in which people that really had very severe Parkinson's disease, as well as dementia, and while a rare Lewy body was detectable in end-stage disease, it was mainly these Lewy neurites that were the dominant pathology. So, I just again want to emphasize that not everything that's bad with respect to the accumulation of these {fibila} proteins occurs in the cell body. Probably 90% or more of the alpha-synuclein deposits will turn out to be in the processes, as is the case for tau.

So, I obviously can't give a tutorial on all things neurodegenerative in a brief 20-minute presentation, but I would like to emphasize this as a schematic that I believe depicts the underlying basis of many of these neurodegenerative diseases, and I'll show

you one slide with a list of some of the more well known ones. And, the basic flaw in sporadic disease, where there's not overproduction or a mutation in an enzyme that's involved in proteolysis, is the conversion of soluble monomers into beta-pleated sheets, oligomers, tetramers, higher molecular weight species that may form spontaneously and, then, revert. They may be eliminated by the chaperone system of the proteasome. But, in those people destined to develop disease, this process goes in this direction, culminating in the formation of amyloid fibers, which also we know, from the vaccine therapy in people, but more clearly demonstrated in animal models, that this process is also reversible. And, the prospect is offered, of course, that if one understood this process well enough, you could intervene at multiple steps along the way to block fibrillization and, perhaps, block it for multiple subunit proteins, Abeta, tau and synuclein, that fibrillize if it was a core mechanism shared by all of these disorders. And, obviously, that it is a question that remains to be answered.

I'll close with this slide showing some of the more common, as well as uncommon, (inaudible) diseases, neurodegenerative disorders that are characterized by brain amyloidoses. In the case of Alzheimer's disease, it's plaque and tangles. But, Lewy body variant of Alzheimer's disease is actually the most common variant of Alzheimer's disease, and that's a triple brain amyloidoses, because you have amyloid formed by synucleins, alpha synuclein and Lewy bodies, tau entangles and Abeta in plaques. And, then, there are single brain amyloidoses dementia with Lewy bodies, multiple system atrophy, tauopathies, trinucleotide repeat diseases and so forth.

So, I will close, and I don't know if I'm going to be able to answer all those wonderful questions. They really – they're great questions that you gave me on the sheet, Neil, but I'm willing to – okay, okay – (inaudible).

Neil Buckholtz, Ph.D., National Institute on Aging

Now, I'd like to ask John Morris to come up and kind of maybe synthesize and open up a discussion section on the first three presentations – the diagnosis issue, the prevalence and incidence issue, and the neuropathology – and perhaps look at the issue of clinical neuropathological correlations, as well.

John C. Morris, M.D., Washington University School of Medicine

I told Susan I have 700 slides that I'm going to – (laughing). Pat, are you able to hear me? Okay. So, I really enjoyed all the presentations this morning. I agree with everything that everyone said, and all the presentations to come, I also agree with everything there, as well. So, for the clinical portion, Ron started us off and I think made a very cogent observation. Alzheimer's disease in 2004 remains a clinical diagnosis. It went through the AAN practice parameter, indicating that the current criteria for dementia and for probable Alzheimer's disease are reliable and valid. Hence, the clinical diagnosis sets the bar pretty high, so new diagnostic tests will have to show how they can add to this already high bar. Ron also indicated that as more information becomes available, at least for the amnestic subtype of MCI, increasingly we're recognizing that many of those individuals already have early Alzheimer's disease.

Denis emphasized how highly prevalent Alzheimer's disease is, and with the increasing aging of the population and the survival of Alzheimer patients, it's going to become even more so in the next 50 years. Denis also made very cogent remarks indicating that when doing prevalence and incidence studies, sometimes the variability comes because of distinctions between normality and disease also can vary. And, so, people make the diagnosis at different stages. And, he also pointed out that translating the high clinical diagnostic capabilities to private practicing physicians often leaves much to be desired. So, when we talk about reliability and validity studies, often these are in more experienced clinicians, and in the real world of private practice, it may not translate nearly so much.

And, then, John just has gone through very nicely all of the different misfolded and aggregated proteins that characterize neurodegenerative disease and pointed out that many of the so-called pathologic hallmarks of Alzheimer's disease are clearly found in many other disorders, as well.

So, let's see, I just want to highlight a few of these points, and then we'll open up the entire period for discussion, just again emphasizing Ron's point about the clinical diagnoses from report, consensus meeting in which Gary Small was the prime convener and lead author, indicating that in most cases, the diagnosis of Alzheimer's disease is clear cut, in most cases. And, in late life, when there is this characteristic picture, usually it is Alzheimer's disease.

And then, as Denis pointed out, the diagnostic threshold for dementia, I think, are in flux. Many years ago, maybe 20 years ago, many people were reluctant to make the diagnosis until what we would consider today moderate disease, and I think people are now more comfortable making the diagnosis in the mild stage. And, as we become more comfortable and as management options increase, I think we'll be making the diagnosis at that stage. But, they're in flux. This is Denis' point, that the actual frequency of Alzheimer's disease in private practice is under recognized. Most primary care physicians have difficulty making the diagnosis, particularly in the mild stage and, consequently, many Alzheimer patients go undetected and, of course, untreated. So, this is a big point that I think we need to talk about as we are, quote, "expert panel," what it's like in the real world, and this is a key point.

By the way, I had thought here at this meeting with CMS to point out that the diagnosis, the clinical diagnosis of Alzheimer's disease, at least in an experienced practitioner's hands, can be, as Ron has indicated, very reliable, very accurate, very sensitive, but it's time consuming, and it takes a lot of time. I think one reason it's under recognized in the community is most physicians don't have the time to do the workup with the family interview and so forth. And, part of that is because the diagnosis process, the clinical diagnostic process is not reimbursed at anywhere near the time involved to do this. It's low-tech, but it's highly accurate. So, as I said, I was going to bring this up at this meeting, but I decided not to mention that – (laughing).

So, we talked about the clinical accuracy of Alzheimer's disease. Here's really an amalgam of two different series, one at Washington University and the other a multi-center study from CERAD, indicating that there are non-Alzheimer conditions that can masquerade as Alzheimer's disease. I think, if my calculations are right, that the overall accuracy of Alzheimer's disease in this combined series was something like 88%. I think in CERAD, it was 85%; in our place, it was 93%. But, there are other entities that can have a clinical picture that looks exactly like Alzheimer's disease, so there are Alzheimer mimics. I still think that the clinical accuracy of 88% is pretty good and, again, I think it sets the bar. We have to look at how well any test is going to provide additive value to that.

There's another study, I don't have a slide from it, from the state of Florida. I think some 382 pathologically diagnosed cases of Alzheimer's disease, and these cases came from community physicians, as well as from state-funded memory disorders clinics, and the positive predictive value of the clinical diagnosis from all of these individuals also was 88% so, again, the clinical diagnosis not perfect, but pretty good.

We talked a little bit about differential diagnosis. I'll just close by showing some data, again, from our center about individuals who meet criteria for *probable* Alzheimer's disease versus those who meet criteria for *possible* Alzheimer's disease. About two-thirds of our cases had the probable AD and about a third had possible AD, roughly comparable on their demographics. And, here are their comorbidities that often precluded the diagnosis of probable Alzheimer's disease and put them into the possible

category. By far, the most common was coexistent depression. But, the possible

contribution of stroke and Parkinson's disease and some, quote, "focally presenting"

conditions also played a role in this. What we did is looked at their subsequent course of

these two groups, probable AD and possible AD, both to the progression and dementia

severity as marked by the clinical dementia rating or to institutionalization or to death.

And, surprisingly, at least in this sample, the presence of the comorbidity did not really

influence the outcome. Once you have Alzheimer's disease, quote. "pure Alzheimer's

disease" or marked by one of these other comorbidities, the clinical course is roughly

the same, and here's the pathological data from that same series. Again, if you look at

Alzheimer's disease from all causes, either with concurrent stroke or with Lewy bodies,

again, the accuracy is 93% for probable AD, and it's 85-86% for possible AD. So, once

again, the clinical criteria are pretty good.

So, those are the points I wanted to highlight and, now, we are to have a

free-ranging discussion of all of these different points. And, I think I'll – well, I guess I

better stand, but I think I don't want to talk anymore. So, I want all – (laughing) – I want

all of you to talk, and I think really we have to talk, but what I'd like to bring up are at

least these things. One is, does it make a difference if we diagnose Alzheimer's disease,

given the fact that there can be Alzheimer mimics, and the histopathologic

characterizations can be quite varied at times. Does it make a difference if we diagnose

Alzheimer's disease? Number two, what about the translation of this clinical method to

the real world of practicing physicians? How are we going to that? And, number three,

what about the changing thresholds? Not everyone labels the disease at different stages.

So, I know there are, you know, so many experience people, so do you want to talk about

any of those issues? Yeah, Kirk does, okay.

Kirk Frey, M.D., University of Michigan Hospital

So, John, the question I wanted to those of you who have access to the clinical

data that we've discussed in terms of the sensitivity and specificity of clinical impression,

the question I would pose is: How good do we do when a patient is seen for the first

time? So, I would draw, by analogy, here back, you know, as I said earlier, 25 years ago

to demyelinating disease where, on first presentation, we didn't do so well. If you saw

the patient three times over a period of two years, we did a heck of a lot better as things

evolved. But, if you wanted to intervene with an immunologically based therapy in MS,

the time to intervene is when the patient has their first attack. So, the question I'd like us

to consider is: How well do we do when the patient is seen for the first time? And I

think this is a reason, in our specialty clinic, we have a four- to six-month wait so that

when a patient gets referred, we have the benefit of evolution of new symptoms.

John C. Morris, M.D., Washington University School of Medicine

So that's a good question. Ron, when you presented the data today, I don't

remember from David's report, how many of those class one and class two studies looked

at the first diagnosis for the reliability and validity?

Ron Peterson, M.D., Ph.D., Mayo Clinic

It's really a combination of both. Most of the studies did follow patients longitudinally, but some of them weren't. And, as you would suspect, the initial diagnosis is not quite as high and that's, you know, the Jobe study in particular was one that was more of a community-based that looked at people initially. So, there clearly is a range of accuracy with regard to the length of follow-up. You know, in research clinics, when you can follow people longitudinally, then your rate goes up quite high.

Christopher Colenda, M.D., Texas A&M University Health Science Center, AAGP

Hi, I'm Chris Colenda. Today, I'm representing AAGP, the geriatric psychiatry group, but my day job is I'm a dean of a medical school. So, I'm in constant conflict, because what's good for my nuclear medicine group is good for the medical school. But, I do have to question, and I'm glad that you talked about accuracy in the positive predictive value, because actually what we're beginning to talk about today is what is the positive predictive value of a test like this in terms of discriminating case from non-case. And, I think the studies that are evident today are on relatively small sample sizes. They are not generalized to the community. They are done in settings where you have an enhanced, perhaps selection bias in terms of the types of cases that hit the clinic, as well as you probably have a selection bias in terms of the types of clinicians who are reading both the images as well as making the clinical diagnosis. So, from a public health perspective and a public policy perspective, I'm not sure that we've gotten enough data yet, well, maybe we'll never have enough data to be able to see whether or not this is generalizable or dissemination can occur throughout the country.

John C. Morris, M.D., Washington University School of Medicine

Yes, so, John Trojanowski.

John Trojanowski, M.D., Ph.D., University of Pennsylvania School of Medicine

I wanted to come back to your point about outcomes, regardless of comorbidities in Alzheimer's disease, and maybe I'm influenced too much by the Newcastle School, where they argue very persuasively that you want to avoid neuroleptics, certainly in dementia with Lewy bodies. And, I don't know if that has been sorted out in Lewy body variant of Alzheimer's versus pure Alzheimer's. And, then, there are also the cost of giving drug to people who have FTD and are not presumably going to benefit, I would guess, from, you know, esterase inhibitors and things like that. So, would you say or would anyone else want to comment on avoiding some medications that could have deleterious effects if you have Lewy body dementia, and giving folks medications that are clearly not going to benefit them if you knew the precise diagnosis more clearly.

John C. Morris, M.D., Washington University School of Medicine

Okay, so a couple of themes have come up. One is: What is the diagnostic accuracy of clinical methods in all comers, people who are not referred to a specialty clinic and already go through some screening, but just show up at any clinic, and what is the diagnostic accuracy of that, particularly the first time the patient is seen? The second is: Does diagnostic accuracy matter? I think that's John's point. In other words, it might matter if a specific medication has deleterious effects for a particular condition, and so

70

how often does that occur? And, the deleterious effects could be both biomedical, or it could be in the cost of treatment that is worthless. So, I know there are, you know, very experienced clinicians here who have the answers to those questions, so can I hear from them? Okay, Claudia. Well, Claudia's actually very good in many ways, but because she also does – she not only runs a clinical corps but also is very versed in epidemiologic studies.

Claudia Kawas, M.D., University of California Irvine

And, I just said under my breath "I'm keeping my mouth shut on this one" – (laughing). Actually, I think I'm supposed to give a presentation on the topic, and I was going to start the presentation saying I really didn't know the answer to these questions. But, we're talking about two different things. Is it valuable now, and is it going to be valuable in the future? And, I think those are two different answers. Right now, I personally am not aware, for example, of any data that says the cholinesterase inhibition doesn't help other dementias. Nobody's ever done the studies in FTD, and many dementias have cholinergic deficits. So, in terms of, you know, doing 100 PET scans to find the five FTD patients to keep those five people off of therapy that may or may not help, I'm not convinced yet that it matters. In the future, though, I think it will, because I think that we're on the threshold of some amazingly specific therapies that are about to come up. And, I think it'll matter a lot when that happens. I mean, in terms of neuroleptics and Lewy body patients, you know, I'm not as convinced as the Newcastle group, only because I've had patients on neuroleptics for years that came to autopsies

with Lewy bodies who didn't have any of those untoward effects. I think there's a specificity to a neuroleptic sensitivity that's interesting but not a sensitivity in Lewy body patients. So, I think I want to stop talking now.

John C. Morris, M.D., Washington University School of Medicine

Do other clinicians have experience with adverse effect of cholinesterase inhibitors in people labeled as Alzheimer's disease but have Lewy bodies or other deficiencies? You know, the few studies of mixed AD, either AD with Lewy bodies or AD with vascular insults or even vascular dementia with cholinesterase inhibitors are all positive. Now, these are all pharmaceutically supported studies and, so, we might think they're biased. But at least, so far, I think as Claudia pointed out, there hasn't been shown in a patient group with dementia that cholinesterase inhibitors are ineffective. I think that's right.

Claudia Kawas, M.D., University of California Irvine

Actually, the pharmaceutical companies, at least in Europe, have conducted at least one or more studies on dementia with Lewy bodies, clinically diagnosed individuals, and vascular dementia and vascular and AD mixed individuals. And, in all three cases, there have been reported positive response to cholinesterase therapy in those groups, maybe because those groups also have AD. I'm not sure why but certainly, in those larger groups, we've already seen that there is some efficacy potentially.

John C. Morris, M.D., Washington University School of Medicine

72

So, to get back to John Trojanowski's query, is it important to really refine the clinical diagnosis in terms of management adverse effects. There's the cost issue and may be ineffective, and we're giving a medicine that's ineffective. But, are there adverse effects? Anybody have experience or want to comment to that? Norman?

Norman Foster, M.D., University of Michigan

I think this discussion is somewhat putting the cart before horse, that is, that appropriate treatments follow from appropriate or accurate diagnoses. So, the examples that I would give are the patients who are put on cholinesterase inhibitors who actually had depression that were not treated appropriately for their depression. I do have some cases of patients with clinically diagnosed frontotemporal dementia who've gotten reportedly worse on cholinesterase inhibitors, but I'm not in a position to decide treatments on the basis of those. However, I would not advocate using cholinesterase inhibitors indiscriminately in everyone who has dementia. Another example would be people who never have the cause of dementia determined and, in fact, they have stroke as the cause of this that becomes evident later. And, appropriate direction of the evaluation and treatment were never directed to the underlying cause of the disease. So, I think it's important that there is great value in accurate diagnosis that goes beyond just the discussion about cholinesterase inhibitors or, indeed, any other treatment for Alzheimer's disease and really is based upon this accuracy.

A question in the interim was asked about what's the value of increasing the accuracy of diagnosis, clinical diagnosis, by something like 10% from 80 to 89%. Well, that's an awful lot of people who are misdiagnosed and may well be mismanaged, even in the examples that I gave of people who we have information on in their entire clinical course. So, there are a lot of dangers of and expenses of misdiagnosis, even when there are small percent changes.

John C. Morris, M.D., Washington University School of Medicine

Kirk?

Kirk Frey, M.D., University of Michigan Hospital

Yeah, I second what Norman said. A couple of anecdotal observations, though. I have seen patients deteriorate even on the so-called "atypical neuroleptics," when the clinical suspicion in our clinic is that they have dementia with Lewy bodies. So, I don't think that this is an infrequent occurrence, and there is certainly the need in patients with more advanced cognitive problems to consider a pharmacological way of suppressing their intrusive behaviors. So, it's not something that is a rare occurrence.

I agree entirely with Norm in terms of needing to make a diagnosis, though. When cross-covering for him in his cognitive disorders clinic, frequently see patients referred from primary care physicians who are already taking cholinesterase inhibitors and, more recently, the glutamate modulating therapies, yet have not had a diagnosis made. And, these patients are financially and medically set back by this. They believe

they have the diagnosis, because somebody's prescribed the medicine. The medication is being paid for largely by them, and it's very difficult to change that impression in these patients when they're already being treated.

John C. Morris, M.D., Washington University School of Medicine

Ron?

Ron Peterson, M.D., Ph.D., Mayo Clinic

Just a quick comment. I mean, I think that all of us in the room who see patients agree that there's value in differential diagnosis when making an accurate diagnosis, even irrespective of treatment implications, just giving the right diagnosis and counseling people. So, as we're talking here about AD versus FTD, I mean, the course of those patients will be quite different. They have different familial implications, the management problems they're going to face down the road, so I think there are some important implications of making the most accurate diagnosis we can. So, I think that's probably well reasoned.

The issue of gaining some percentage of accuracy in the diagnosis versus the expense is a much more dicey issue, and that I think needs further discussion.

John C. Morris, M.D., Washington University School of Medicine

So, there are two ways that's a dicey issue. One is for people who come to Mayo Clinic or to the University of Michigan, the value of any adjunct procedure, in this case

PET, to improve the accuracy 1%, 5%, 10% versus the practicing physician, the primary

care physician. What do we know about that? So, you know, some reports suggest that

50% of all demented patients seen in private practices have no dementia diagnosis, go

unrecognized; 50%, as high as that. So, would PET be useful to help practicing primary

care physicians in the diagnostic, and is it cost effective?

Male Voice

        Yeah, I think you have – (inaudible/off microphone) – you asked the right

question. This is a matter of public policy, because most patients are initially seen by a

primary care physician, not in specialty clinics, and think of this as a toolbox. What is

the repertoire of tools that a primary care physician has in his or her box to make a

diagnosis, a complex diagnosis? I would submit that we don't know enough about the

generalizability or the dissemination of these types of technologies into real-world

practice to be able to ask and answer the question yet.

John C. Morris, M.D., Washington University School of Medicine

        Does anyone know enough? Mike Weiner knows enough.

Michael W. Weiner, M.D., VA Medical Center

        I know enough to say that there's no data on that – (laughing). The imaging work

over at the Alzheimer's Association recommended the value of the use of PET scanning

after a thorough and complete diagnostic evaluation by an individual who's experienced

in doing that. So, I think that what follows, if there are a lot of physicians out there who are dealing with people who have problems with dementia, the solution is not for them to go out and order PET scans. The solution is for them to get better trained in how to evaluate the patients or to refer patients to physicians who are capable of doing it.

John C. Morris, M.D., Washington University School of Medicine

Or maybe reimburse – no, no, no, I'm not going to – (laughing) – no, I'm not going there. I'm not going there.

Michael W. Weiner, M.D., VA Medical Center

The other topic, though, I thought it was – up until now, the discussion has largely been on the issue of differential diagnosis between one form of dementia and another. Earlier, Eric addressed the issue of the value of PET in the early detection of Alzheimer's disease prior to the onset of dementia. I thought he showed some interesting, although very preliminary, data showing how uncertain the value of PET was for the early identification of dementia. And, since this is a use that PET seems to be quite widely touted for, I think it would be useful to have some discussion about what is the practical use of PET scanning in people with mild cognitive complaints or family histories or people who're clearly not demented but are being worked up.

John C. Morris, M.D., Washington University School of Medicine

And, Eric, you were talking in large part to people enriched with ApoE4 positivity, but will you remind us all again about studies of PET in individuals who meet MCI criteria and the value of PET in predicting the group who will go on to have Alzheimer's disease? Can you remind us about that?

<u>Eric Reiman, M.D., Good Samaritan Regional Medical Center, Phoenix, Arizona</u>

I'll come over. But, I would like to focus more of this discussion on the diagnosis and clinical pathology correlations, because we're going to have a separate discussion on PET in a few minutes. That said, I referred to small studies that have looked at the predictive value of PET in predicting weight of conversion to probable AD, and the studies are small and promising but certainly not definitive. And, they really call for more studies, not for clinical use at this time.

<u>John C. Morris, M.D., Washington University School of Medicine</u>

And, I think you said it really still is in the research tool stage.

<u>Eric Reiman, M.D., Good Samaritan Regional Medical Center, Phoenix, Arizona</u>

Exactly.

<u>John C. Morris, M.D., Washington University School of Medicine</u>

Okay. Norman?

<u>Norman Foster, M.D., University of Michigan</u>

Well, just to point out that patients with frontotemporal dementia may not have an abnormal Mini Mental State examination or may not meet DSM for criteria for dementia, because they don't have a prominent memory deficit. So, in that case, there may be a better role for PET to help identify those. But, again, it's in those who have symptoms that would suggest highly the possibility of frontotemporal dementia.

<u>John Trojanowski, M.D., Ph.D., University of Pennsylvania School of Medicine</u>

On the clinical, pathological correlation theme, I don't see a speaker or topic related to biomarkers, but are there competing technologies that are coming on strong that may – well, not competing, complementary, I guess you could say also – that may be informative. We know that there's CSF, tau and Abeta, and there are isoprostanes and, then, single lipids that David Holzman is working on. We're putting together a study of FTD where it seems as though there is an informative protocol or a profile that would distinguish AD from FTD, wherein the tau levels are not elevated for some strange reason as they are in the CSF of AD patients. Abeta levels are close to normal, but you wouldn't confuse the individual with someone who's normal cognitively clearly have a neurological disease. But, the profile may be useful, the CSF marker profile may be useful in moving you toward the diagnosis of FTD. This is, you know, preliminary data, because the paper hasn't even been written yet, but I'm certain other people doing these kinds of studies, (inaudible) and others, that may enable us to have biomarkers that would help increase our certainty of diagnosis.

John C. Morris, M.D., Washington University School of Medicine

And, I would go so far to say that if we ever do develop a biomarker diagnostic, it probably will not be a single biomarker. It probably will be a combination maybe of cerebrospinal fluid plus imaging plus something else, but some biochemical assays with some either structural or functional measure of the brain I would imagine is the way we're going to go.

Well, let's come back to Denis Evans' point about the variability – where does cognitive normality end, and where does disease begin, and how can PET help us with that? So, who wants to tackle that one? Maybe that first one is too tough to tackle – (laughing) – when does cognitive normality end and disease begin? And, let me throw out one other point, John, I don't think brought this out, but it's clear that the more we study brains of putatively non-demented older people, the more we see Alzheimer pathology. Now, does that mean that these people died before they were going to clinically express dementia? Or is this simply aging phenomenon, they never would have become demented. And, how would that affect imaging studies, functional imaging PET studies? If there's Alzheimer pathology in normal people, would that translate into regional hypometabolism, and would we call these people AD profile, and what would we do with them if they're cognitively normal?

Neil Buckholtz, Ph.D., National Institute on Aging

Well, let me ask you, John. Maybe you could just talk about some of the data that have come out of Wash. U. about that issue in terms of maybe even the relative contribution of plaques and tangles in normal aging versus pathology to kind of set this up a little.

John C. Morris, M.D., Washington University School of Medicine

You know, contrary to many people's assumptions, we don't have all the answers at Wash. U. – (laughing). We do not. But, what we find pathologically, in cognitively normal people, particularly once they reach 75 or older, is about a quarter will have sufficient plaque and tangle pathology to meet neuropathologic diagnosis for Alzheimer's disease. So, Ron and our group and other centers are involved in a NAC-sponsored study that is showing in additional sites the very similar percentage. At least 25% of normal older adults have neuropathologic Alzheimer's disease, as defined by current criteria. So, the issues are, I think, are these individuals examples of presymptomatic Alzheimer's disease? They harbor the disease. It's getting to a point where, ultimately, enough brain damage/brain failure will occur, and they will ultimately dement if they live to that age. That's one thing. Second is, they have neuropathologic Alzheimer's disease, but they have protective factors so that the disease is not expressed. For some reason, they have protective factors. And, the third would be that the neuropathologic markers that we use to diagnose Alzheimer's disease can be part of normal aging. That's another possibility. You know, we favor the first interpretation, but we don't know that. We don't know that these people will ultimately develop Alzheimer's disease, although I think that's, you

know, again, talking about research uses of PET, that's clearly a very appropriate application. If people who are in a normal population show PET features of Alzheimer's disease, they could be followed longitudinally to see if they are at increased risk of developing symptomatic Alzheimer's disease. So that's why I asked. In cognitively normal people, could we use PET to highlight those people at very high risk of becoming demented? That was my query.

Neil Buckholtz, Ph.D., National Institute on Aging

And, actually, I'll take my prerogative –

John C. Morris, M.D., Washington University School of Medicine

Okay.

Neil Buckholtz, Ph.D., National Institute on Aging

– to hold that question. If you will ask that question when we come back to the PET discussion and then go on to Keith Johnson's presentation, because we just have one presentation, then we're going to come back to the PET discussion and we can fold this into the PET discussion if that's okay.

So, I'd like to go ahead and ask Keith Johnson to come up. Keith Johnson from Mass General, who's going to talk about SPECT and, then, we're going to come back to a general discussion on PET and SPECT after that. And, Keith, while you're waiting, if you could just talk about any conflicts.

Keith A. Johnson, M.D., Massachusetts General Hospital

Yes, sure. I have no conflicts. Susan asked me to comment on some of the postmortem findings and some of the data that's come out over the years, and I'll do that briefly. So, I will briefly cover some of the autopsy data that's come out, look at some findings in early diagnosis, and conclude with a little bit of differential diagnosis. There have been several series; I just want to point to three that have studied postmortem confirmation of SPECT diagnosis. These are series that, as you can see, date from the mid to late 90s and, of course, because this field is so rapidly changing, technically, all of these studies were done on cameras that are basically now nowhere to be found, because the improvements have occurred, and these are just not cameras that are in use anymore.

What this data tells me is that while there is general agreement with the kind of numbers we've seen before, sensitivity in the mid to high 80s, specificity much lower, the caveats that have been emphasized before have to be kept in mind, it seems to me. These do not account for the prior probabilities of people coming into the studies, as Ron Peterson pointed out earlier this morning, which will of course have an impact on the numbers that come out of the comparison. And a related issue, of course, is that many of these patients have been selected because they are relatively non-confusing clinically and, so, the numbers here presented don't necessarily reflect the kind of performance one might expect in a sort of true-to-life world situation in which all comers are taken.

The Optima data set, which is represented first here, was reanalyzed by Bill Jacobs and colleagues, and a similar study was presented earlier. But, if you look at the additional contribution of the imaging – for some reason, the pointer doesn't seem to be here, sorry. So, without any imaging, the correspondence between the clinical diagnosis of probable AD and the postmortem diagnosis of definite AD was about 84%. If you add a positive SPECT, that climbs to 92. If you add a negative SPECT, that goes down to 70. And, there's corresponding information for much smaller numbers of cases in which there's a possible AD. So, I think this indicates that there is a history of this kind of study, which is in accord with the kind of history we've seen from the PET data, with the caveats that I've mentioned, both technical and in terms of the study design.

Here is a more recent example of a SPECT study, in comparison, the same patient at the same level, and I just want to make some general comments about the comparability of the two. I think you can appreciate, even non-imagers can readily appreciate that there is a greater definition. We use the term "resolution." There is also greater sensitivity, which translates to more narrow variance in the statistics that are used to calculate quantitative assessments of these data. So, while there's no question that PET technology involves higher spatial resolution and sensitivity, it's my view that these are likely important in many circumstances, but they're not as critical when abnormalities are extensive. And, I think the other side of this is that, you know, given the context of 2004, SPECT is currently much more widely available. However, the number of PET

facilities is increasing quite dramatically, in large part because of the applications in oncology that have become far more widespread since 2001.

I just want to briefly review some of the findings of SPECT imaging in early diagnosis. You can actually look at, in Alzheimer's disease, the number of abnormal voxels and, in this study, comparing a PET camera to an older SPECT camera, the correspondence is pretty good. This is the number of abnormal regions in Alzheimer's disease in one technology compared to the other. As was mentioned earlier, one of the challenges in trying to identify the earliest changes in Alzheimer's disease is to see if carriers of mutants which produce the disease demonstrate abnormalities in their imaging prior to the development of clinical symptoms. We were able to look at a family who carried the presenilin-1 mutation from Colombia, and you can see, similar to the example of the single case shown earlier that, as a group, the people who are asymptomatic have abnormalities compared to their kindred – in the same kindred who do not have those abnormalities.

The other way of approaching this, of course, which probably has more direct interest, is to look at individuals who are at risk for Alzheimer's disease, because they have complaints of memory impairment, so-called "questionable Alzheimer's disease," and this is data from Dr. Albert's longitudinal study of such subjects. And, areas of abnormal profusion at baseline correspond to those subjects who eventually convert to Alzheimer's disease. And, after a longer period of follow-up, we've been able to identify some differences between subjects who are at risk because they're questionable and

decline, from those who, in addition to declining, actually convert so that it looks promising at this point that we might be able to identify in a baseline image an abnormality which would indicate that the person is actually going to not only decline, but actually convert to Alzheimer's disease. And, this is some data that we'll be presenting at the World Alzheimer's Congress this summer.

So, finally, just a brief mention of some information that I think has not perhaps, at least from the imaging side, been exemplified quite as much as it perhaps should be, and that is look at dementia with Lewy bodies. FTD metabolism, here represented as Z-scores on the surface image, are abnormal compared to normals and are different than Alzheimer's patients, so that a finding which was first reported from the group in Michigan really seems to be holding up, that in addition to the association abnormalities in the parietal lobe, you see association cortex abnormalities in occipital lobe in patients who have a simultaneous reduction in the dopamine transporter in the striatum, in comparison to Alzheimer's cases, who have no such reduction in the dopamine transporter. Just to emphasize that the direction of using this kind of imaging for differential diagnosis and some of the key breakpoints in the pathology, I think has a very promising future, and we are just starting really to be able to tease these things out. So, I think I'll stop there and turn this back over. (Audience/inaudible). No, these are all clinical diagnosis, all clinical diagnosis. That's right. (Audience/inaudible). Right, by DLB, I mean consensus criteria, yeah.

Neil Buckholtz, Ph.D., National Institute on Aging

Thanks, Keith. At this time, I'd like to ask Mony DeLeon to come up. He's going

to be the person who's going to lead the discussion on PET and SPECT, and Mony has a

few comments to begin with.

Mony DeLeon, Ed.D, NYU School of Medicine

Thank you, Dr. Molchan and Dr. Buckholtz, for your invitation to come here. I

have no conflicts of interest that relates to this presentation. Actually, while they're

getting the slides set up, let me offer that today being, you know, the day just before

Passover, where a seder is really part of the tradition, the word "seder" translates to

"order," and perhaps it's not all that out of context that we're trying to get our own house

in order today with respect to some of these PET-related matters.

I have three major goals, three major points to bring out. The first has to do with

the early and its specific diagnosis for Alzheimer's disease. The question for discussion

is: Where are we? Certainly, this has been echoing through the morning. The second

point I'd like to address has to do with how PET data are actually acquired. What is a

PET scan that we're going to be using? And, the third issue is really with respect to the

analysis and the interpretation of the PET scan.

So, I'd like to start by just discussing very briefly how good are these prediction

studies. Well, there are a number of studies that are out there. There are cross-sectional

studies in the amount of about 250 of them. There are 37 longitudinal studies, and this is

plotted as a function of five-year intervals on the x axis. Overall, you sort of get a bottom

line very quickly. There are five prediction studies that have tracked the conversion from MCI to AD. What they have in common is a reference of decliners versus non-decliners. These are not specificity studies, so to speak. What they have in common is pretty good sensitivity, running at about 85%, and also pretty good specificity but not decline, running at about 80%. So, with respect to the prediction of decline from these research samples, PET is doing a pretty decent job.

Just for a moment, an issue was raised a little while ago about the use of PET potentially in unaffected people with normal memory functions, and this is a result from a study that we recently published, showing how if you can sample the entorhinal cortex, one can detect evidence for change that, over time, is consistent with the overall future diagnosis of MCI over an interval of about three years. This is an FDG-PET scan, again with sensitivities and specificities in the 80% range.

But, the real heart of the matter is what is the deferential diagnosis question. That's where the issue hits the road. With respect to deferential diagnosis, there are about 20 studies. There's less than 20 studies that are out there over this same period, from 1984 to the present. If you look across these studies, you see that there are a number of features of the PET scan that are considered to be relevant for the differential diagnosis with respect to {dimensions} of Lewy bodies. The occipital lobe has been highlighted as an area of maximal information relative to Alzheimer's disease. In other words, occipital metabolic reductions are considered to be useful. In the case of normal pressure hydrocephalus, global changes have been recognized. In the case of

frontotemporal dementia, as we've said and heard earlier from Norman and from others, at the frontal lobe, it's particularly informative. For clinical depression in Alzheimer's disease, there's very little out there. And, for any of these entities that I just mentioned with respect to MCI, there's no data. That's where the red zero goes for, so the differential is early. With respect to MCI, I could not find any evidence for.

Now, how good are these differential studies? Well, here's a different story. Here you see that the sensitivity for Alzheimer's disease is really quite good. In other words, recognizing Alzheimer's disease in mixed cohorts is quite good. However, recognizing not Alzheimer's disease or specificity is not so good. And, as we've seen today from Dr. Frey and Dr. Foster, consistent with the data on this chart, the specificity runs less than 70%. So, we have a good ability to detect the Alzheimer's disease in the brain, but we have far less a capacity for recognizing the other degenerative diseases using the FDG-PET scan.

Here's an example. This is like some of the detail of a paper that was referred to several times this morning. This is the Silverman paper from UCLA. What you see on this chart, these are cases now that have had the pathological validation of their dementias using FDG-PET. The red arrows point to the frontotemporal dementia cohort, which is an {N of seven}, and the dementia with Lewy body, which is an {N of six}. The accuracies for those two small groups were quite variable. For example, five out of the seven of the FTD cases were diagnosed as not having Alzheimer's disease, which is a good thing. But, only two out of the six dementia with Lewy body cases had a diagnosis

89

of not Alzheimer's disease, which is not such a good thing. If you actually summarize

the specificity evidence for the cases on this chart, and you take out the cases where no

what they refer to as "neurodegenerative dementia" was present or cases without

histology, you're left with 23 cases. And, of the 23 cases, 14 of them are correctly

identified as not having Alzheimer's disease, putting you back into a specificity range of

about 60%. So, I would argue that the specificity question, which is paramount in this

arena for getting these tools out into the clinics, one has to reckon with accuracies that are

in the 60 to 70% range.

So, now, another issue that I think is worthy of discussion, does PET add to other

modalities? Well, quick surveys through the literature shows that PET does in fact add

to the recognition of normal MCI and AD when compared to MRI, but this is not a

differential diagnosis. This is again the normal versus AD or the normal versus MCI

kind of question; in other words, the research question, not the application question. PET

adds to SPECT in the differential, but the gains are small, and these studies have largely

been done in Europe and in Japan.

Well, does PET add to MRI? In a study that Susan Desanti did, and Susan's

actually here today, she found that when focusing on co-registered PET scans with MRI,

that PET does add to the MRI when looking at the hippocampus, for example, but not in

the case of normal MCI. PET is equivalent to MRI. It adds to MCI in a comparison with

AD, and it adds to the normal versus AD. In other words, the sensitivity of PET boosts

the differential over and above the MRI, or it's equal to the MRI. And, this is for a relatively small region, where PET is at somewhat at a disadvantage.

The final point on the diagnosis side has to do with these e4 carriers. Well, we know about these patterns of change, but we really don't know about the risk associated with the pattern. That was a very poignant point that was made this morning by Eric. Does the e4 genotype actually confound the diagnosis? Does it make it more difficult to make a diagnosis? Here are some data, and forgive me if you don't like the way these images look, because it was just a quick cut-and-paste job. But, these are data from Eric Reiman on the left here in his very nice study showing how e4 carriers, when they're young, show this pattern of metabolic reductions in the parietal lobe and in the temporal lobe. Gary Small's data shows an effect in older normals. In other word, e4 carriers, when they're older, show this metabolic signature, if you will. And, the data here by Lisa Masconi from the group in Florence, and Lisa's actually here at the meeting, as well, also shows this effect with respect to MCI. This is (inaudible) data. Lisa's data is not published yet. So MCI patients, carrying or not carrying the e4 genotype, will also show this pattern. Well, if you're looking at patterns and if pattern recognition is your issue, then unless you understand the risk associated with the pattern, you've got a dilemma. Okay.

The second major point that I think is worthy of discussion is the PET scan itself. What is the PET scan that we want to bring out? Well, a cardinal question in PET scan is: Do we need quantitation? Should a PET scan have numbers, or it should it be a count

91

image? Should it be a relative image? Is it important to know what the actual absolute

metabolism of a particular region is? We have a problem across many studies that are

in the literature, because there's a lack of uniformity. We have a problem with the

denominator. What is the reference? If we don't have the absolute value–

(End of Tape 2, Side 1)

– they require blood sampling. If we don't have the absolute values, and we're

working with ratios or we're working with patterns that are visualized, well, how are the

images going to be standardized? A related question, which was brought up earlier in

some way, is a single scan necessary for an early diagnosis when working with a pattern,

and we'll have to see the pattern change in order to be confident. In the case of an e4

carrier, perhaps you would. If you had the signatures at baseline, then are you going to

get sick or not is really unclear.

A final point for discussion has to do with the interpretation itself. There are

many ways to look at PET scans. The literature is replete with them, and all of us in this

room have contributed our own favorite strategies. And, some of them are quite

effective, and some of them have been replicated. And, there's some very good tools out

there. However, we need to decide how we're actually going to do this. Are we going to

eyeball the scans, as Norman Foster did in his study? Are we going to do ratio

examinations of the data, sampling with or without an MRI? Are we going to use a

voxel-based analysis, either the SSP we heard about or the SPM analyses that are

available to us? And, if so, what do we lose when we do that? The voxel-based analyses don't typically give us information about the medial temporal lobes. MRI data certainly show that the medial temporal lobes are highly relevant, but the voxel-based morphologies are extremely reliable. The computer solves the problem all the time the same way. The computer's very good at that, but is it giving us the information that we really need, especially early on. I mean, these are questions. These are not conclusions, I hope you realize.

And, this middle point here, B, if you're looking at a scan, do we need reference examples? In other words, at a national level, do we publicize this is what a mild AD looks like? This is what a typical FTD is. This is what levels of pathology look like. And, are there going to be norm values, like a clinical lab? How are we going to reference this material? How are we going to provide the uniformity of the observation?

And, finally, do we need an MRI scan? It seems that, from what I heard so far, that an MRI scan is part of the routine clinical workup. If that's the case, wouldn't it be an intelligent thing to do to have that MRI scan collected in such a way that it would actually be useful for the interpretation of the PET scan, as opposed to two independent acquisitions that themselves are not necessarily talking to each other.

So, where do we go from here? And these are really my views of some of the questions, instead of a restatement. I think the PET sensitivity is good for the detection of Alzheimer's disease, and the question remains and certainly for this group to discuss

about the current utility with respect to differential diagnosis. Do we need standards and acquisition of interpretation, the point I just made? Are we ready for the introduction of PET now, because that's certainly going to be an issue worthy of discussion? And, then, the final question is really: Are we waiting for a newer imaging initiative and, if so, what are we waiting for? Are we going to have what we need, post-initiative, to address some of these questions? So, that's really as far as I wanted to go, and I hope I didn't take too much time from the discussion.

Male Voice

Mony, since you brought up the issue of how the initiative that's going to be funded by the NIA would relate to this, this is a topic that keeps coming up. I think that there really is very little relationship between the goals of the initiative as the way they're stated in the RFA that was sent out and these kinds of diagnostic issues. The initiative will use very carefully selected patient populations, and the data from the initiative would barely touch on this. What the initiative could do, though, is set certain standards for acquisition and processing, which you talked about a little bit in your talk, that might be used to help create more uniform standards for approaching the diagnostic questions. That is a potential outcome of it, but that's not certainly part of the goals of the initiative the way the NIA has written them in the RFP.

Male Voice

(Inaudible/off microphone) – I do think there'll be some informative data that comes out in looking at how baseline measurements, the combination of PET and MRI measurements, will predict rate of conversion from MCI to AD. That will help to drive those additional, more generalizable studies that need to be done at that time.

Susan Resnick, Ph.D., National Institute on Aging

I'm Susan Resnick from the NIA Intramural Program, and the question I wanted to ask, which relates to the point earlier about pathology in people who are still clinically normal, is: What information do we have with PET in terms of differential sensitivity and specificity in the oldest old? So, is it equally effective when we're talking about people 80 and older?

Mony DeLeon, Ed.D, NYU School of Medicine

Do we have any – I'm sorry, Claudia. Do you have any information yet?

Claudia Kawas, M.D., University of California Irvine

(Inaudible/off microphone) – I think the short answer is almost no relationship. We have very little data for sure, but what we do have makes it clear to me that it's not going to be what we thought. I mean, I've always been impressed with – 20 years ago, our diagnostic accuracy for pathology was only about 60% in all the published literature. And then, more recently as you see, you know, the clinicians are able to get 85 or 90 or 95%. So, in the first 12 brains that have come to autopsy out of my study, I think I'm

running about 20% accuracy. We have had patients with florid dementias, Mini Mentals of two, that have come to the autopsy with normal brains. We have had at least three people who were demented who came to autopsy, and although they had a little amyloid and a little tangles, nothing that reached the level that we would call pathologic AD. And, we've had the opposite. So it's fairly clear to me that all bets are off when you get into at least the 90s. Of some interest, I have a living subject right now is an e4/4 and is normal, 95.

Male Voice

So, Mony, you raised a number of issues, and I wrote down a couple of them just to give my take. I think the first thing, though, to point out at the outset, what we've been doing so far is talking about the use of imaging alone as a diagnostic categorization. And, I think that the studies that have looked at this have done the same. So, you know, we're appropriately reporting what scientists have published. The critical issue here, though, to migrate this into a clinical paradigm or not, is to decide how this plays out with conjoint knowledge of the history and physical examination, obligate ancillary laboratory studies that all of us would do in the setting of suspected dementia. And, for instance, the problem that you raised with the Silverman paper, I agree. It looks like in the non-AD pathologies, there is a less than optimal ability to characterize and classify them properly. But, some of those are, to some degree, red herrings. So, for instance, a dementia with Lewy bodies would be very difficult to distinguish from Alzheimer's disease, with the exception of maybe the occipital lobes. And, if you looked at

Silverman's categorization procedures, they did not identify nor did they ask the interpreters to identify additional occipital involvement. So, it should have turned out that that was the way.

In our own studies, we would never have mistaken the demyelinating disease cases for AD. We would have the history of repeated attacks and the MR showing demyelination. And, similarly in the Silverman case, there are a number of disease entities that you would not clinically confuse for AD. So, I think we need to take those values, you know, as they are, but to think about how they would migrate into the clinic where this would be like any other diagnostic test. It could increase or decrease your diagnostic confidence but should not be considered in isolation.

Male Voice

So you would argue that additional work is required in order to actually apply the most current understandings of the expression of pathology on a PET scan, so that those criteria could then be seen as how they play out as either independently reflecting that pathology against postmortem, or complementing the clinical diagnosis and seeing if there's any added value. So, there's something missing there. What's missing is the current view (inaudible).

Male Voice

So, I'm not sure that to be accepted for clinical use, it has to have that additional qualification established in a prospective, independent, you know, blinded, randomized

fashion. But, the work that Norm Foster showed, for instance, does indicate how the added value of the imaging data refines clinical diagnosis in a way that suggests it is useful and does net positive contribute to making the proper clinical assessment (inaudible).

Mony DeLeon, Ed.D, NYU School of Medicine

I realize this shouldn't be a forum for debate. I haven't seen Norm's paper, which I learned is in preparation; I guess it's been submitted. But, there was the issue of how the clinical examinations were, in fact, categorized. The review was actually see the patients or see the materials from the patients. And, so, the base, it's another sort of red herring type issue, where the best of what could be done hadn't been applied in that design. So, I'm not sure that that's great evidence in support of the added value of PET. Please, go on.

Male Voice

Yeah, well, I just want to make two other quick comments. First, with regard to are standards necessary and does one require quantification, I think the answer there is no. The largest studies that you've cited and the meta-analyses that people have made are largely related to the pattern of abnormality rather than quantitative data. And, I think the notion that there's no single approach to quantifying and interpreting these studies actually speaks to the notion that one does not have to have an externally imposed analytical strategy or framework in order to achieve these kinds of results. So,

everybody's different; I agree. But, there's great consistency in terms of the performance of FDG in the same kind of clinical context. And, I'm not so sure that one needs the MR. The data that I've seen on MRI rely on interval change and loss of structural volume. The impact of structural imaging, as I understand it, in the initial evaluation of dementia relates to the exclusion of alternative diagnoses. And, in many clinics, MR is used but, in others, high-quality CT may substitute, as well.

Mony DeLeon, Ed.D, NYU School of Medicine

Would anybody like to pick up on those points?

Male Voice

Yeah, I just wanted to say that there's a paper – I didn't have a chance to discuss this in my presentation, this paper by Roger {Higben} in statistics medicine this year, in which we looked at these same cases and tried a number of statistical algorithms to determine whether any was preferable or more accurate than visual interpretation, and the answer was no. The second is that when we looked at inter-rater reliability for evaluating specific regions, it was lower than for overall assessments. So, coming up with a strategy which would just look at some combination or some algorithm from individual areas is significantly lower in reliability than an overall interpretation based upon a pattern. So, some of these questions you raised have been pretty well answered.

Male Voice

If I may return to Susan's question and ask the PET folks in the room who have done a lot of studies in this area, how good are the data on normal elderly subjects period? So, Susan said in the oldest old. But, even in 70s and then you get in the 80s, when you take John's data that he was talking about, the frequency of Alzheimer changes in clinically normal people, where Claudia was saying about how difficult it is to make diagnoses in the oldest old, I guess I'd be surprised if PET were really able to distinguish normal aging from some form of pathology when you get into these relatively unselected populations of older people. But, I could be wrong. Are the data that good in the normative populations, in normal subjects, or are they rather, again, select subpopulations?

Mony DeLeon, Ed.D, NYU School of Medicine

So, does anybody want to speak to that (inaudible)? Is there anyone?

Susan Resnick, Ph.D., National Institute on Aging

Our data, though, is O15-water, so it's blood flow, not FDG. And, we do have data, and we do see, you know, fairly pronounced – well, we see some longitudinal changes over the four-year period that we've looked at so far, but we also see cross-sectional age effects in some of the regions that we're talking about. And, that's actually what I was asking was whether or not there is this added benefit of PET in people who are 80 and older. I don't think there's enough data out there to address that right now.

Male Voice

Our PET data on aging go up to age 80, and we see these preferential reductions in prefrontal and anterior cingulate cortex, a more modest reduction, age-related reduction in the posterior cingulate. I agree that it would be nice to have more data, but I believe that when you control for age in comparisons of patients versus controls, you'll end up seeing the same pattern that we see in younger patients.

Male Voice

I would confirm that actually, yeah. If you look historically back at {Drower's} work and years ago in the 80s and our work, {Bud Insejagus} and {Buddinger}, all those studies showed a preferential involvement to the frontal lobe with a relative sparing of the temporal lobe. But, of course, these are done on somewhat lower resolution instruments, and whether or not some of the finesse that we have today can be applied to those aging studies and will yield a different answer remains, as Susan puts, still unknown. Our tools have gotten better.

Male Voice

Thanks, (inaudible).

Male Voice

Two quick points. One is to the question we barely heard the issue about when in the disease the patient is seen, questions as to how good are the clinical diagnosis, going

back to Ron's, John Morris' point, and what does the PET add incrementally? You've

seen a lot of numbers, and most of them are from published reports. I would suggest that

when the patient is beyond the mild stage of disease, moderate and more advanced, the

diagnosis is far more apparent, a good clinical diagnosis, not at the hands of somebody

has not worked up the patient and has prematurely put someone on cholinesterase

inhibitors, but from individuals who worked up the case well. It gets much better. You

have a much better sense of what you're dealing with. You have some history of

progression. You have behavioral manifestations. You have sort of a complex clinical

web to make a diagnosis from. Consequently, in cases like that, PET is probably

somewhat marginal, as opposed to maybe early on in the cases when it's somewhat more

complicated and unclear, based on the clinical symptomatology.

The second issue is if PET scan cost is just $100, you wouldn't be here discussing

this. Okay? The reality is we've had numbers somewhere between $2,000 and $3,000

put up. And, at some point, perhaps after Dr. Matchar's presentation, sometime in the

afternoon, I would hope we have more of an integration of cost effectiveness in this

process. So, that's the backdrop against what we're having this entire discussion. It's

not nearly incremental utility in the diagnosis of AD; it's incremental utility and can we

afford that at this stage, and that's what's driving the debate in many ways.

Mony DeLeon, Ed.D, NYU School of Medicine

Claudia, did you want –

Claudia Kawas, M.D., University of California Irvine

(Inaudible/off microphone) But, everybody today keeps showing us data that's group data, and this is true. I mean, I know that if you take a group of demented people and a group of normal people, whether they're 90 or they're 60, that you will find group differences in their studies. But, what I haven't heard is how confident are you in single person prediction, which is really what we're trying to decide today when we're talking about diagnosis. That is, if I give you a scan or a PET scan or a SPECT or something on a 92-year-old, what can you tell me about that individual, rather than the group of individuals whose data we look at in research? I mean, can you tell me a diagnosis and that it's normal or not?

Male Voice

I can quickly respond to that. I mean, the largest series that you've seen today, the data from Silverman, the data that I showed from our prospective trial, the data that Dr. Foster showed and others, are individual subject analyses. Now, the more subtle aspects of pre-clinical evolution that Dr. Reiman showed are, indeed, summation across multiple individuals, and I think this is why Eric is appropriately conservative in not arguing for diagnostic use in those clinical indications. So, at least from my perspective, I would not advocate the use of FDG or any other screening test like this to decide whether dementia is present or not in someone with no symptoms and no signs. So, I'm not arguing that we should take everybody at 90 years of age, scan them and decide what

their problem is on the basis of the scan. On the other hand, in an appropriately selective clinical setting, I think there is clear value to the incremental data that we're discussing.

Male Voice

I think that the answer to your question has to do with the receiver operating curve characteristics of this particular test, and none of the data that's been presented today has shown ROC analyses apply to predictive value. You know, the ideal ROC analysis would be with a square curve with the angle to the top left of the curve. My sense, given the specificity and sensitivity data here, is that the curve is not going to be a 45-degree curve, but it's going to be significantly skewed to the right and downward, so that it's going to be very hard to do individual predictions based upon any level of knowledge that we have on receiver operating curve characteristics.

Male Voice

Well, the sensitivities and specificities seem to be uniformly good with respect to the diagnosis of AD or MCI against control, so there you would have very nice ROC curves. When you extend that analysis to the differential diagnosis, where your sensitivities are on the order of about 80% and your specificities are hovering around 65%, then I would agree that that curve is not going to be as pretty.

Male Voice

But, you have – (inaudible/off microphone).

<u>Male Voice</u>

But, you can infer the curve from the sensitivity and the specificity.

(Inaudible mixed voices and female voice/off microphone).

<u>Male Voice</u>

The sensitivity and specificity should be relatively independent of that.

<u>Female Voice</u>

But, it's not.

<u>Male Voice</u>

No, it should be and it is, as long as you have appropriate confidence intervals. So, if your N is enough to get a reasonable confidence estimate, sensitivity and specificity should not be affected. What should be affected by the overall prevalence of disease is the accuracy.

<u>Neil Buckholtz, Ph.D., National Institute on Aging</u>

Okay, on that note, I would like to introduce Dr. David Matchar from Duke University. Dr. Matchar's presentation is somewhat out of order, because he has to leave early. But, again, this is getting into the very important topic of neuroimaging studies viewed from a clinical decision-making perspective, and Dr. Matchar will go ahead now.

And, then, he's going to have to leave, but one of his colleagues is going to be here for this afternoon's discussion.

David Matchar, M.D.,Duke University Medical Center

Good morning. So, I'm standing between you and lunch, which is always better than standing between you and having had lunch. Actually, this is a pretty good time for me to get to talking this morning, if any time is a good time at all. I think that Dr. Kulasingam and I are here today, and Dr. Kulasingam will be here later after I leave, and the area that we work in in our center is not specifically in Alzheimer's disease, although Alzheimer's disease has been an area that we've worked in for the last couple of years, specifically in working with CMS, with Medicare, but in generally looking at innovative technologies like PET scanning and various other technologies for diagnosis and treatment and asking the question: How does this really relate to clinical practice and clinical policymaking? Now, even though, as I mentioned, we've done work with Medicare, and I'm going to talk this morning a little bit about – a lot about the work that we've done on a technology assessment on behalf of a coverage decision that CMS was in the process of making, that we didn't make, that they made. And, I'm going to go through this in order not to talk specifically about the technology assessment and certainly not to talk particularly about the coverage decision, but mostly to ask the question: How does this kind of assessment speak to the question of what research does one need in order to answer the clinically important questions? It's not that I'm backpedaling away from what CMS decided. I'm simply trying to make the distinction

106

between the agenda for today, which is what I understand my charge to be, which is to better understand what kind of research agenda NIA might be interested in promoting.

Well, in going back to this technology assessment, I'm going to go back to the original CMS-sponsored assessment, which had the following aim. It was to assess the benefits of FDG-PET scanning in patients with dementia, with mild cognitive impairment, and in asymptomatic patients with a family history of Alzheimer's disease, subsequent to the standard evaluation as described in the American Academy of Neurology guidelines. Now, this was I guess a couple of years ago, and this is part of a process that Medicare has been really promoting over the last few years, which is an open, evidence-based process for making decisions about coverage. Our group is an evidence-based practice center that happens to have a contract with the Agency for Health Care Policy and Research and, indirectly, our commission to work for other Federal agencies, in this case, CMS. And, I guess that's the sum and substance of my disclaimer, which is that I get part of my salary, the contract that ultimately are fed by CMS dollars.

Now, the original technology assessment was a literature review, but what was innovative I think about the way CMS chose to approach this question, again, it was in response to a specific request for coverage for PET scanning in Alzheimer's disease. What was innovative was their request for a decision model that would help provide an understanding of the decisional context in which testing would be done, and that's why I

107

think that this discussion perhaps is a useful exercise following the discussion that we just had.

Now, again, in the spirit of its being publicly available, open and above board, the coverage decision and also all of the parent or the source documents that went into the evaluation for CMS that was then subsequently incorporated in some way in their coverage decision, those are available on the Internet, so feel free to go check it out. It's all in the public domain.

Now, the first thing that we did when we looked at this question was: Is it possible to make the direct inference that testing leads to better outcomes, where "better outcomes" are defined specifically in terms of things that people care about? And, we defined, and this was in conjunction with CMS and other clinicians, that the three areas that seemed to be important are the one that – the delayed progression, which is one of the responses potentially to treatment; decreased mortality, although there doesn't seem to be clear evidence that such a thing occurs in response to delayed progression, but if it were to occur, people would care about it; and, then, there is something we'll call the "value of knowing," which is the value of being given a true diagnosis. That's exclusive of its utility in decision making for therapy.

Well, unfortunately, as I say, there's no randomized control trials, for example, that would allow you to say, "Okay, we randomized this group of patients to receive PET scans, this group of people not to receive PET scans, and let's see how the thing plays

out, how clinical decisions are made, how people are treated, and what their quality of life, what their disability status is, and various other outcomes, what their mortality is." So, what we approached this was and this was, again, somewhat innovative from the perspective of a lot of evidence-based reviews, in that some people would simply stop and say, "There is no evidence that if you were to use testing, that it would lead to better outcomes." And, that's what Medicare was asking for, better outcomes in terms of the things that Medicare recipients will care about. And, so they would stop but, in fact, they said, "Okay, it's okay for us to look at a causal chain that would link testing to these things that people care about." So, if we were able to establish that the elements of the chain were strong, then the inference that testing led to those outcomes could indirectly be deemed to be true. So, the way by which testing leads to positive outcomes is through the mechanism of achieving true positives, primarily.

Now, there's also this other possibility that you would have true negatives and having a true result leads to the value of knowing, okay. And, again, this is exclusive of any affect that it would have on treatment decision-making. Now, of course, there are downsides that need to be considered, which is the adverse effects of treatment which we also examined. And, then, this is just the flipside of the true positive/true negative aspect of the sensitivity specificity two-by-two table is the false negatives and false positives. And, when these occur, you're unable to achieve any of these good things; you just miss the boat entirely.

Now, the important point is that true positives and the definition of "true

positives" are kind of key to any further analysis. How does one define a true positive?

Now, one could use a histopathologic diagnosis, but if you're looking in terms of effect

of treatment, what seems most appropriate and what our clinical experts agreed was

what's most appropriate in that case is the clinical diagnosis that was used in establishing

who was and was not an acceptable subject for treatment trials. So, the reference

standard actually is clinical diagnosis.

Now, you mentioned earlier these ROC curves. What we did as part of the study

was – one component of the study was to do a comprehensive review and metanalysis.

This is actually a meta-analytic receiver operating characteristic curve and, then, a mean

estimate or a best guess estimate and confidence box for estimates of sensitivity and

specificity for PET scanning. I'll just point out to you that there was a wide variety of

studies clearly that were included here, some of which seemed to have nearly perfect

operating characteristics. But, on balance, I think everyone seems willing to stipulate

that the sensitivity of PET scanning is somewhere in the range of 85 to 95%, and

specificity is going to be lower.

Now, you see, that's another interesting point, which is that this is an aggregation

of all of the studies, which includes both the distinction between people with AD and

people without AD, and they could have been normal individuals, and also looking at

people at different levels of severity. So, the tendency was that sensitivity and specificity

would be better when we were trying to look at where your spectrum was at the extremes,

an AD patient versus a normal patient. Where this issue comes up of sensitivity and specificity is not a stable characteristic of a test, okay? It is not a stable characteristic of a test. It depends highly on the distribution of clinical characteristics in the population that you're looking at. So, in that case, sensitivity and specificity would look artificially and in a biased way too high, and it also includes patients who had, as I say, higher and lower levels of severity, and the operating characteristics with lower levels of severity tended to be not as good.

But, then, getting back really to the punch line, not the exact numbers, but let's say for the sake of discussion we had three options that were being considered, given that *a priori*, we figure that an individual who has mild dementia is believed by the clinician who would at that point treat them, to have a 56% probability of having Alzheimer's disease. Or, let's say that they would be willing to diagnose Alzheimer's disease, but they would only be 56% likely to be correct, okay, consistent with the notion that doctors in the community are a bit dim – (laughing) – okay? That would be one option. The other one would be to use PET in that circumstance to discriminate between those people who you would then subsequently treat as positive. And, then, this just as a comparison would be what if we just left all those patients alone. So, if you look at it in terms of just the way the true positives and so on fall out, in this circumstance, given that 56% are positive and are deemed true positives because you treat them as though they do, in fact, have the disease, you give them medication, and so you have a 56% correct diagnosis

rate. In fact, since everybody is treated as though they have a disease, you don't have

any false negatives, okay.

Now, under the PET scheme, it falls out differently, such that you have many

fewer false positives. We go from 44% false positives to 6% false positives, and you

have a few more false negatives. And in balance, net, you have greater accuracy, and I

think that that's a point that's been made by some of the advocates and researchers in the

area of PET scanning is that you do have a higher net accuracy rate. So, again, I think

everyone's probably willing to stipulate that point.

And, then, the question is: So what then? What does it all mean? And, one of the

problems is, given that you don't have a clinical trial that says what happens when you

randomize people to PET or no PET, and you have to make the inference that if you're

going to decide on treatment on the basis of the PET result, how is that going to play out

in subsequent disability and death? What we did was develop the model, which is not

dissimilar from other models. In fact, Dr. Gazelle later is going to show you a model that

I think is quite similar to this one, in which we basically simulate a population. So, we

can generate a synthetic population of individuals based on C-RAD data and, then,

making the assumption that that treatment has a specific level of efficacy in terms of

reducing the rate of progression and also has a certain risk of adverse events. Okay?

And, so, you just basically use this calculation model to play out the result of this

decision – the effect of treatment and see what happens. And, I'm not going to go into

the painful details because, again, you can go to the website and find all the details about

how the model was constructed and so on. But, the punch line is that even though the

overall accuracy of the PET, that a treatment decision conditional on PET was highest, in

fact, the life expectancy, quality adjusted life expectancy in what we call the severe

dementia-free life expectancy, was actually a twinge shorter on average, if you followed

a PET scanning strategy. And, so, a primary conclusion is that PET could improve

overall accuracy compared to the accuracy of an exam based on AAN guidelines.

However, treatment based on AAN recommended examination leads to better health

outcomes than treatment based on PET results.

So, how do we make sense of this apparent discordance between these two

analyses? Well, just recall that the net accuracy with PET is better because there are

many fewer false positives and a few more false negatives. So, when you take them into

account, the many fewer false positives add up to a net greater accuracy. However, false

positives are not the same thing as false negatives. Now, exactly how different they are is

a point of speculation and an empirically testable question. However, okay, so

incorrectly treating, which is a false positive, is not as bad, I would argue, is not as bad as

incorrectly not treating, which would be a false negative, because incorrectly treating, the

treatment is relatively benign. It may be beneficial, as has been pointed out earlier, even

if the patient doesn't have AD. We don't know that apparently. There has been some

speculation that that's true. Incorrectly not treating, however, the patient loses the benefit

of treatment. And, if one argues the treatment is effective, and I haven't heard yet

anyone say that treatment is not effective, then this is actually a worse thing to have

happen. So, even though, say, 7% fewer – the overall accuracy is higher, those seven out

of 100 people who are not treated as a consequence of having a false negative study

actually, on balance, lead to a worse average outcome.

Now, on the positive side, when would testing be preferred in terms of this

measure of aggregate, of life expectancy, quality adjusted life expectancy, severe-

dementia free life expectancy, presumably things people care about? If a new treatment

becomes available that is not only more effective than current treatment but is associated

with a risk of severe adverse effects? Okay, now if it's as effective and has more severe

adverse effects, then, of course you wouldn't want it. So, if it's going to be more

dangerous, it's got to be a whole heck of a lot more effective. And, in this case, this is a

diagnostic issue. Here PET scanning could become, or any kind of imaging technology

would be beneficial because you're avoiding the false positives. You've increased the

value of the thing that the test is best able to accomplish.

The second area in which testing could be preferred is if the results have

demonstrable benefits beyond informing the choice of therapies, what I was calling

earlier the "value of knowing." And so here, for example, it becomes a prognostic

question. Your ability to adjust and to plan has been argued to be a major benefit of

accurate diagnosis.

The third area where testing would be preferred is if testing could be demonstrated

to be a better reference standard than an examination based on AAN guidelines. As I

mentioned earlier, it was baked in the cake. We didn't have to do a fancy evaluation, a fancy Markov model, in order to show that PET scanning wasn't going to be net better, because the reference standard for evaluation for treatment is the AAN guidelines. But if, in fact, you could say that, in looking at this response to therapy question, if you could say that compared to a standard examination, testing is better, or testing better distinguishes who respond to therapy or who will have adverse affects to therapy, if you could do that, then you could make testing worthwhile.

So, several months ago, there was a new request for coverage to CMS, and we were asked to not only look at not only PET, which is what we were asked to look at before, but also to look at SPECT, 3T MR looking at volumetric measures, functional MR and MRF. And, here, we were going to focus, then, on those three questions now that we felt that we had some insight into what was important to look for in research, we thought we would focus then not only on diagnosis, which is the question which everybody tends to focus on, prognosis in response to therapy.

So, I'm going to go through this real quickly, because this is a work in progress. We have a draft that we're working on for CMS, and we're actually going to be looking for additional potential references from some of you, okay? And, just to show you that, this is a distribution of technologies that we discovered. Now, this is of 328 at least ,and I think this number is actually higher, we reviewed 328 abstracts in full text form and, of those, only 28 were deemed to provide information of any of the sorts that we were asking for. When we looked at the clinical questions that these 21 studies looked at, they

all related to clinical impact in terms of accuracy of diagnosis or prognosis. None of

them were looking at response to treatment, and absolutely none of them were looking at

the value of knowing. And, when we looked at quality measures, and these are relatively

standard quantities that are evaluated when we look at clinical diagnostic tests, we

discovered the reference standard was defined in 100% of studies, because that was

required to be included. But, consecutive or random enrollment was relatively

uncommon. There were no trials, even though we looked for them. There was very little

attempt to match or adjust for imbalances between or clinical features that we knew about

in advance. Only two, or 10%, looked at providing an *a priori* cut point.

So, now, just I've got three quick slides looking at our research agenda in service

to decision making, so now we're beginning with the end in mind. We have a decision,

let's work backwards and figure out what we need to know in order to make that

decision. This is where the literature is about technical feasibility and optimization,

about 90% of the literature is about this, maybe 9% is about this, and less than 1% is

about that, diagnostic thinking, therapeutic choice, patient outcome. These are categories

that CMS has been using now in order to assess the literature, the scientific evidence as

they consider their coverage decision. So, it would be good to look at this particular

article if you're interested in studying these issues.

I think that this issue has been raised multiple times, and I'm just going to raise it

one more time, which is: What's the incremental benefit? The odds, if you're going to

look at the long odds of having Alzheimer's disease, it's going to be a function of clinical

116

characteristics, conventional testing, and whatever your new test is. You've got to take all of these things into account, and it's not really fair to anyone just to treat the new test as though it's something in isolation.

And then, study design. This is all pretty much Mom and apple pie among the clinical research community, and it's only when you go back to the literature and you discover how few of these criteria are being satisfied, that you recognize that when studies are done in future, they ought to very explicitly use this as a checklist. And, I would suggest that given the fact that we're talking about potentially maybe a billion dollars per year in utilization of a new technology, that, you know, a few tens of millions of dollars for a clinical trial might not be such a bad idea. That's it. Thanks.

Neil Buckholtz, Ph.D., National Institute on Aging

Since Dr. Matchar has to leave, I thought we'd take five minutes to ask any questions that you have of him in the methodology.

Male Voice

So, just a question, so are you advocating that all patients over the age of 65 should take a cholinesterase inhibitor? I mean, this is the implication of your analysis.

David Matchar, M.D.,Duke University Medical Center

No, it's not. It's not in any the implication of my analysis. I just suggest you read it. It's actually – what it says is that if a clinician feels that this patient is likely to have

117

Alzheimer's disease and has only 56% chance, 50% likely to be correct, and that was just based on your conjecture that clinicians are not particularly good at this. And, so, I accepted that poor performance.

Male Voice

I was curious about how you would propose a study that takes into the account an ancillary benefit of PET and that is, in the real world situation, I'll tell you what I see occasionally anecdotally. A primary care physician who's referred a patient for PET has not done the laboratory dementia workup, and he calls and asks: "Is this indicated?" And, then, we review what's involved in a comprehensive dementia evaluation, why are they ordering the scan. They go back. They do the studies. They may or may not be referred to PET at that point. But, the fact that PET came up brought them to increase their sensitivity, whether or not they get the PET. So, if you want to talk about the added value, how do you somehow get those subjects and those physicians into a study in a realistic way? It seems to me, in a study, you're already biasing yourself to those individuals, those physicians who have recognized the problem.

David Matchar, M.D.,Duke University Medical Center

Well, that's why you would do it as a trial. That's one of the reasons, one of the explanations for doing it as a trial. I mean, if what you're saying is that by having a PET available and having a PET result available, you can make people do things that they should have been doing anyway, I might suggest or someone else might suggest that you

could use the same amount of resources to actually figure out how to get them to do the right thing or to teach them to do the right thing or to refer them to somebody else who would do the right thing.

Male Voice

(Inaudible/off microphone) – how to do the right thing (inaudible) evaluate the decision of (inaudible) the person of whether to get the PET (inaudible) to use the PET results.

David Matchar, M.D.,Duke University Medical Center

I don't know the answer to that question. I may not understand the question. That may be the problem.

Male Voice

In your simulation, did you say you used the CERAD data set?

David Matchar, M.D.,Duke University Medical Center

That's right.

Male Voice

Does that influence the outcome insofar as people who got into the CERAD study were certainly not random samples of people, like Eric saying, who might come to a general practitioner. People who got into the CERAD study were very finely refined and

had, you know, a certain probability of having AD and certain characteristics. So, did that at all – could that have played a role in the outcome?

David Matchar, M.D.,Duke University Medical Center

No. I mean, the reason it wouldn't, it might have changed the magnitude of the difference to the extent that people might more rapidly or more slowly be moving through this process, this, you know, this life process. But, as long as the impact of treatment was to diminish the rate of progression from state to state, then the results would have been the same.

Male Voice

You mentioned the "B" word, billion. Dan Moerman, a former colleague of the University of Michigan group, and I have done a series of studies looking at cost of Alzheimer's care in terms of direct, total direct, and indirect costs. And, clearly, the most influential component of cost of care is behavior and changes in behavior. When we're talking about a billion, I'd sure like to see some of that billion go into treatment for behavior because, there, you might realize the most cost benefit. In the cost benefit analysis, you might realize the best gains there as opposed to up-front diagnostic testing.

David Matchar, M.D.,Duke University Medical Center

Well, actually, we're modifying the model at this point now to look at issues of behavior and how modifying behavior might play into the outcomes and cost of care for Alzheimer's disease. But, that was not included here.

Neil Buckholtz, Ph.D., National Institute on Aging

Thank you very much. I think we're ready for lunch. Again, let me make an announcement, any Federal employees need to give their money to the contractor. Non-Federal employees are free to enjoy the lunch – (laughing).

(Lunch break)

Neil Buckholtz, Ph.D., National Institute on Aging

– Michael Weiner from UCSF, who's going to be talking about MRI in vascular dementia versus Alzheimer's disease.

Michael W. Weiner, M.D., VA Medical Center

Thank you very much. In preparation for this talk, I just wanted to acknowledge a number of collaborators and colleagues. For the last 10 years, Helena Chui has had an NIA program project grant on subcortical ischemic vascular dementia, and Bill Jagust participates in that; he's at UC Berkeley now. And Charlie DeCarli recently moved to the Bay Area; he's now at UC Davis. And, all of these people have contributed information for this talk, and also {Philip Shelton}, who's in The Netherlands, is also very interested in the diagnosis of vascular disease, of vascular dementia, and he also sent

121

me some slides. I also am going to talk about a consensus statement concerning the

Medicare reimbursement for PET, but I'm scheduled to talk about that at a later time this

afternoon, so I'll just skip over that.

So, what is the role of MRI in the diagnosis of dementia? Certainly, MRI may

detect mass lesions or other diseases which cause cognitive impairment. Although CT is

frequently used, MRI has greater sensitivity than CTs to detect brain disease, so many

people, especially in academic centers, tend to favor the use of MRI, even though it's

somewhat more expensive. I must say, I don't think that I've ever seen any comparison

of CT versus MRI in the utility to really rule out disease in an Alzheimer's workup. MRI

is thought to have some additional value in the diagnostic process, in addition to ruling

out other disease. But, the data on pathological correlations in MRI for a differential

diagnosis is minimal.

What about the value of MRI to detect Alzheimer's disease? Alzheimer's disease

is clearly associated with atrophy of the hippocampus, the entorhinal cortex, and the

whole brain and the parietal lobes. Visual inspection of the MRI tends to help confirm a

diagnostic impression. So, if a clinician has worked a patient up, feels the patient has

Alzheimer's disease, looks at the scan and sees small hippocampi, that's reassuring to the

clinician, perhaps falsely reassuring. Atrophy tends to favor a diagnosis of Alzheimer's

disease or MCI. Lack of atrophy may raise concerns, but there is really not that much

data actually to support this in the diagnostic setting. There are huge amounts of data on

the hippocampal volumes in AD and MCI and so forth. What there's much less data on

is how visual inspection of the MRI and looking at the median temporal lobe structures, how valuable that is in making or assisting in the diagnosis of patients and separating Alzheimer's from controls, as well as separating Alzheimer's from other dementing diseases. For example, given these two scans and one sees all this atrophy here, one's going to easily say that this patient's mostly likely to have Alzheimer's disease; in this patient's, this is a healthy control. That kind of clear difference is easy. But, let's say you have a situation like this. Let's say, here, you have a 79-year-old woman, and this is one hippocampus, looks fairly plump for her age. And, here's somebody else with a smaller hippocampus. Well, would you tend to say that this person has Alzheimer's disease, and this is the control? Well, in this case, you'd be right. This is a healthy elderly control, and this is an Alzheimer's subject. Here's another example, though. Here's a patient with a hippocampus that's bigger than this hippocampus. Which is the patient that has Alzheimer's disease? Well, I selected this just to show how much overlap there is. Here's an elderly control, which had no cognitive impairment, and this patient was diagnosed clinically with Alzheimer's disease, even though this person's hippocampus is bigger than this. So, we all know this. There's a lot of overlap among subjects, which is why you cannot rule *in* a diagnosis of Alzheimer's disease with a structural MRI.

What about the value of MRI in differential diagnosis? Well, in terms of frontotemporal dementia, Norman has already told us quite a bit about his work using PET. There is really no data that I know of on the value of MRI for the differential

diagnosis of frontotemporal dementia versus AD. Bruce Miller at UCSF sees a lot of patients with frontotemporal dementia, and he allowed me to quote him now. He says that he believes that MRI has great value for the diagnosis of frontotemporal dementia and, in his opinion but without any data to support it, he feels that MRI is at least as good as nuclear scans for the separation of frontotemporal dementia from Alzheimer's disease. But, we don't have any data to support that statement.

There's no data on MRI in Lewy body dementia that I know of, so what about vascular disease and vascular dementia? Well, MRI clearly is the best way to detect cerebrovascular disease. It's the imaging modality with the greatest sensitivity to detect the effects of cerebrovascular disease. Well, this is just an example of an FTD patient versus a control. So, MRI is very sensitive to detect the effects of cerebrovascular disease, with especially T2-weighted MRI. We see cortical strokes, subcortical strokes, lacunar infarcts, and white matter disease, much of which is associated with cerebrovascular disease. MRI cannot diagnosis vascular dementia; vascular dementia is a clinical diagnosis. But, imaging is important in the diagnosis of vascular dementia, because you need to have evidence of severe vascular disease to make the diagnosis.

So, exactly what are the criteria for vascular dementia? This is the most generally accepted criteria. You have to have dementia, which is a clinical diagnosis. You have to have evidence for cerebrovascular disease, based on a clinical exam and imaging. And, one needs to demonstrate some relationship between the cerebrovascular disease and the

dementia on a clinical basis. The type of cerebrovascular disease in these criteria are

large vessel strokes in these territories and/or extensive small vessel disease.

Now, what is the actual hard data that we have correlating evidence with

cerebrovascular disease seen on MRI with pathological data? There's some out there in

the literature. This is the study that Bill Jagust is currently writing up, based on data from

our program project grant. This is 43 subjects whose mean age adjusted was 83. There

was approximately 1.8 years between the MR scan and the time of death. And, of these

43 subjects were seven normals, 10 MCIs and 26 demented subjects.

What these graphs are showing is the imaging data on the Y axis and the

pathology data on the X axis, showing a correlation between white matter signal

hyperintensities, white matter lesions, measured by image segmentation, with the

incomplete infarction in the white matter shown and graded quantitatively on pathology,

so quite a good correlation there. The extent of atherosclerosis also correlates quite well

with white matter lesions. In contrast, the number of tangles or the Braak stage is not

correlated at all with white matter lesions, as you might expect, and the CERAD score,

plaques, also does not correlate well with white matter lesions. But, the Braak stage does

correlate well with the hippocampal volume. So, in general, kind of the data has shown

what you would expect, that is, that cerebrovascular disease, the pathology correlates

with white matter lesions seen on MRI or as the extent of Alzheimer's disease plaques

and tangles shows on pathology correlates with hippocampal atrophy. And, I think Mony

DeLeon's group has also reported correlations between hippocampal size, clinically

hippocampal size and neuronal density at autopsy, as well as plaques and tangles at autopsy.

This is another slide from Charlie DeCarli. I think this is from the cerebrovascular health study and simply where they correlated the cognitive state of subjects with the extent of white matter signal hyperintensities, showing that people with a lot of white matter disease have greater impairments. But, we all know that even normal subjects with no cognitive impairments can have quite extensive white matter lesions. So, once again, there's a great deal of overlap here.

So, in summary, MRI measures as cerebrovascular injury and neurodegeneration are good surrogate markers of pathology. It should be emphasized that no one claims that MRI can diagnose vascular dementia. Vascular dementia is a clinical diagnosis. MRI is highly sensitive to detect effects of cerebrovascular disease, especially infarction atrophy. It's a vital tool for the diagnosis of vascular dementia and mixed dementia. It's useful to determine the extent of cardiovascular disease in aging, Alzheimer's and other dementias.

Since we are talking about the use of FDG-PET in MRI for the diagnosis of Alzheimer's disease, I thought it'd be interesting to just give you a little update or some new information about the use of perfusion MRI. We know that FDG-PET detects changes in Alzheimer's, MCI and other dementias, and we also know that it's widely accepted that cerebral blood flow and glucose metabolism are closely related. Therefore,

perfusion of MRI of the brain should detect the changes similar to PET. And, if one could develop a useful perfusion MRI scan, incorporated as part of a structural MRI scan, it's possible that one could obtain the same kind of information that you get from PET scanning as this part of an MRI protocol. And, a lot of investigators in our laboratory are really focused on trying to accomplish this.

So, I'm just very briefly going to show you some recent results on 29 controls, 17 AD, 16 MCI subjects. There's ages and there are the MMSE scores on the subjects. The data has all been analyzed in SPM format, so we're showing group differences. This shows the changes in blood flow, comparing the normal controls to the Alzheimer's disease. This shows the changes in structure, that is, gray matter atrophy, Alzheimer's versus controls. So, you see in the Alzheimer's patients, we're seeing a parietal lobe hypoperfusion, especially in the region of the posterior cingulate, as you might expect from PET. And, now, here's our patients with – cognitive impaired, non-demented subjects, kind of MCI patients with a CDR of 0.5, compared to control, and we see statistically significant reductions of perfusion in the posterior cingulate, just like Eric and others have shown with PET.

So, there needs to be much more work done in this area, but the perfusion MRI detects changes in Alzheimer's disease, MCI. Perfusion MRI can easily be added to a structural MRI scan. The scans take about 12 minutes. Perfusion MRI may provide similar diagnostic information to FDG-PET without the use of isotopes at low cost so, certainly, this is an area where a lot of future research is indicated.

The next part of the talk is the consensus statement about Medicare reimbursement for PET, but I think we're just going to put that off until later today. So, I will go to the next talk.

Neil Buckholtz, Ph.D., National Institute on Aging

Thanks, Mike. As we did for this morning, we'll have the presentations and then we'll have a longer session for the discussion. The next presentation, there's a slight change, will be in the order, Marilyn Albert will be talking about MRI in mild cognitive impairment and AD, cross-sectional and pathological correlations.

(End of Tape 2, Side 2)

Marilyn Albert, Ph.D., Johns Hopkins School of Medicine

– potential to use these kinds of measures in clinical practice. I should apologize at the outset. I'm going to present data from a lot of people in this room. I've often been accused of being an example of fool rushing in where angels fear to tread – (laughing) – and I hope all of you will forgive me if I don't entirely do justice to the data that you have.

So, first of all, what do we know about MRI associations with respect to the diagnosis of AD and MCI, and I'm going to talk about identifying these individuals cross-sectionally and, then, using data from baseline to predict subsequent development of disease. But, the data from baseline are cross-sectional data and, of course, not

longitudinal data, and Ron's going to be talking about longitudinal data in a few moments.

I should just begin by saying that we all kind of take for granted that basically what we're trying to do is capture the evolution of pathology in the AD brain. This is a very nice slide that was developed by Juan Troncoso at Johns Hopkins, showing that what we think is going on is the evolution of all of these pathologies across time, and what we're trying to do is see if, with the imaging modalities we have available, we can get an indirect measure of these pathologies. Obviously, we would all like something that's much more direct, and you've already heard some reference to various new forms of PET imaging and biomarkers that might be able to do that.

As you heard from Mike, when MRI was first used with respect to Alzheimer's disease, the main focus was on looking at regions in the medial temporal lobe because of what we know about the evolution of pathology in the hippocampus and now, more recently, in the entorhinal cortex. Mike showed you a slide very similar to this, where you can see these regions visually, but a lot of the work that was done initially, looking at visual ratings, demonstrated that they were really not highly accurate, and it was very difficult to teach multiple clinicians to do visual ratings in a standardized way. So, what evolved over time was, in research settings, taking MRIs, having people with neuropathological skill outlining regions of interest and, then, taking the volumes of those measurements in order to try to see if we could identify cases and controls.

The initial work, as I mentioned, that went on was in looking at the hippocampus. Subsequently, it was recognized that pathology began perhaps even earlier in the entorhinal cortex. So, the early work in Alzheimer's disease, comparing mild AD cases to controls, was done largely looking at the hippocampus and, more recently, there've been a whole lot of different studies comparing AD cases to controls, looking at both the hippocampus and entorhinal cortex. And, this is just an example of some of them.

These are data actually from Mike Weiner and his colleagues, looking at the accuracy, the sensitivity and the specificity for the hippocampus alone, the hippocampus plus the entorhinal cortex, and the hippocampus and a measure of gray matter volume from segmenting the entire image. And, I think what you can see is that as you add these measures to one another, you have increasing accuracy and increasing sensitivity and specificity for the discrimination of Alzheimer's disease versus normal controls. But, as you've already heard this morning at some length, the clinical diagnosis for Alzheimer's disease is quite accurate. The cases that were included in this study and in many other studies that I'm going to mention were already carefully screened. And, so, the question we would have to ask, if we wanted to think about using this kind measure in clinical practice, is: Does it provide any added value?

The accuracy of diagnosis goes down a bit when you start looking at AD versus MCI. Obviously, if we are correct that the MCI cases, particularly the form with memory predominant disorders of MCI, are likely to develop AD, this is a much more difficult discrimination, because presumably people with MCI have a lot of the underlying

130

pathology as well as the cases with AD. And, you see that the accuracy goes down, as does the sensitivity and specificity but, again, you see that as you add measures to one another, you do get increasing accuracy.

In this same paper by Mike Weiner and his colleagues, what was shown very nicely is that although these measures are additive, they do in fact correlate very highly with one another. And, so, you are getting some added benefit from measurement but, in fact, there is a good deal of redundancy with these particular two measures that I've been talking about.

These are data from Leyla De Toledo-Morrill and Brad Dickerson and their colleagues in Chicago, and they found very similar findings. If you just look here at the volume of the entorhinal cortex in very mild AD cases compared to controls, you see a very significant difference and, likewise, a difference with the MCI cases, but less striking and the same thing here. Oh, that was the hippocampus. This is for the entorhinal cortex.

Turning to trying to use these data for prediction turns out to be actually much more relevant to the clinical question that all of us are increasingly struggling with on a daily basis. Most of us feel as if we can make a clinical diagnosis of Alzheimer's disease with some degree of accuracy, but most of us who are seeing patients clinically and are trying to make a prediction about the evolution of Alzheimer's disease in people who are not demented feel less certain, considerably less certain about that prediction. We might

think that, over time, it's very likely for an individual to progress to dementia, but it's very hard to be certain exactly how long that will take. And, all of us, I think, who see these patients clinically have a much greater degree of uncertainty, so this is an area in which imaging could potentially play a role.

The experimental studies that have looked at this have had a very similar format of design. Subjects are evaluated longitudinally over time to see what happens to them clinically, and the data that was collected at baseline, in this case an MRI scan, is used to try to predict the evolution of disease over time. So, the data I'm going to be talking about don't involve the repetition of the MRI scan. That's what Ron is going to be discussing in a few moments, but rather the annual evaluation of the subjects is repeated, and the data and the subjects are sorted in terms of what happens to them over time. But, the data are actually cross-sectional data obtained at one point in time. There are several studies that suggest that MRI measures in general can tell you that people are at greater risk if they have reduced volume of the MRI. These are data from Michael Grundman and the ADCS cooperative study that is led by Leon {Saul}, and this is a study that is currently just winding up. But, what they did was to look at the volume of the hippocampus and to look at how rapidly people progress to dementia over time. And, I think it's very clear from looking at this that people who had lower volumes of the hippocampal formation were at much greater likelihood of converting to Alzheimer's disease over time than the people whose volumes were substantially larger.

These are data from Cliff Jack and Ron Peterson at the Mayo Clinic, and you can see how similar they are. So, again, what they did was to take people and follow them over time to look at the volume of the hippocampus and its relationship to the likely evolution of disease. And, again, you see that people whose hippocampal volumes were very high were much less likely to progress to a clinical diagnosis of dementia than people who had much smaller volumes.

Brad Dickerson and Leyla De Toledo-Morrell have very similar findings, but they've been trying to look at the differential use of the entorhinal cortex versus the hippocampus. And, what they argued was that in their study, the volume of the entorhinal cortex, which was a much better predictor of conversion to subsequent development of disease than the volume of the hippocampus. And, what they found was that they had an odds ratio that was statistically significant, suggesting that there was a 7% increase in likelihood for every 10 units of entorhinal cortex volume. But, in their hands, the hippocampus was actually not useful in this increased prediction of individual progression to AD.

In Boston, with my colleagues, Ron Killiany and Raoul {Defacan} and the group that Keith mentioned earlier, where we've been looking at the evolution of Alzheimer's disease in a similar setting, we found that the entorhinal cortex had a gradual and progressive decline in volume over time that was quite striking, even among people with relatively mild symptoms. Again, this is cross-sectional data, so these are data obtained at baseline but sorting people on the basis of how they progressed over time. But, at least

in our hands, the hippocampus was not nearly as striking. However, what we found was that the entorhinal cortex alone was exceedingly useful in discriminating individuals, this is sort of relevant to the question that Claudia asked earlier, how well you could predict what was going to happen to individuals, and the entorhinal cortex alone was very helpful in discriminating groups of individuals, except in the clinical comparison of greatest importance, discriminating the people who were subsequently going to go on to develop dementia within about three or four years versus the people who had impairments, the people we called "questionable" but who had not progressed to this diagnosis over time.

And, so, what we ended up doing was looking at more than one region and, in fact, regions outside of the medial temporal lobe. We looked at the entorhinal cortex. We looked at the banks of the superotemporal focus, because our colleagues, Brad Hyman and Theresa Gomez-Isla had found that this also showed pathology very early in the course of disease. We looked at the cingulate because of some findings from Keith that the middle cingulate was important in predicting evolution of disease and, then, we looked at a lot of ventricular regions. And, what we found was that if we wanted to be really accurate in the clinical discrimination of importance, this difference between the people who didn't progress within three to four years versus those who did, that three measures in combination enabled us to make that prediction about individuals over time, and this just shows you what these regions look like.

So, this is what the accuracy and sensitivity and specificity was when you combine those regions together and you ask the question could we predict on an individual basis the likelihood of progressing to a dementia over time. And, again, you can see that the greatest accuracy is in comparing the groups at the two extremes, the controls versus the converters, and a much lower accuracy was in comparing the people who are more similar to one another, both of whom who had cognitive problems at baseline and were not demented, one of whom went on get a diagnosis of disease over time.

So, just to sort of summarize what I've been saying, from this part of the data, I think that there is little question based on the data that I have mentioned, as well as lots of other findings in the field, that measures of the hippocampus and the entorhinal cortex are associated with increased risk of developing AD. If one looks at what measures are best for predicting evolution of disease over time, I think there's still a considerable amount of controversy. It's much less clear which measures are best. It's much less clear whether or not you can use an individual measure or even combinations of measures to predict evolution of disease. I think it's likely that this debate is because of differences in the way the regions of interest are measured to begin with, but it's probably more a result of who the subjects are, exactly what the clinical criteria are for the subjects who've been involved. So far, the data suggest that if you combine measures together, you're much more likely to be accurate. That's certainly what our data suggest, and I think that Mike's findings would warrant a similar conclusion.

So, now, what about the relationship of these measures to cognition? Fortunately, in this sphere, there is complete unanimity. I have some data to just very briefly show you that says I think pretty convincingly that cognitive measures that are related to memory correlate very consistently and very highly with measures of the hippocampus or the entorhinal cortex, as you would want them to be if these measures were meaningful and related to the disease you were trying to follow. So, these are data again from Mike, looking at the relationship of the Mini Mental to measures of the hippocampus and the entorhinal cortex, and you can see that there is a highly significant correlation. These are data from Ron Peterson and Cliff Jack, looking at the correlation of the Cued Selective Reminding Test and the volume of the hippocampus. These are data from Mike Grundman from the ADCS trial, looking at the correlation between the hippocampus and the NYU Paragraph Delayed Recall. And, these are all data from Mike Grundman from the ADCS trial, looking at the correlation between the hippocampus and performance on Delayed Recall of the {ADA COG}. And these are data from Boston, from my colleagues in Boston, again, seeing the same exact thing. And the one thing that I would like to emphasize is that there are also data, both from Mike Grundman and from ourselves, indicating that it's not that everything correlates with everything else. Mike in the recent article that he published argued very convincingly that memory tests correlated with the hippocampus but non-memory tests and we find a very similar thing. We get correlations of executive function tests, for example, with other regions in the brain, in this particular instance, the anterior cingulate but not with the hippocampus or the entorhinal cortex.

So, now let me turn briefly to the issue of correlations with pathology. Mike mentioned a little bit about this already. There are really two studies and then the one that he just mentioned, which I guess is in preparation, one looking at the relationship between the volume of various regions of interest and pathologies and looking at the volume of the region when it's measured on pathological tissue, and the other looking at the relationship with Braak staging.

These are data from Mony DeLeon and his colleagues at NYU where they outlined the parahippocampal gyrus and the hippocampus, the subicular area, and they did that on MRI. And, then, they looked at the same individuals when they had autopsy tissue available. And, they looked at correlations between the autopsy tissue and the MRI volume, and you can see that the correlations were really quite astonishing, considering that there is brain shrinkage when you fix the brain, and you're taking somebody afterwards. The fact that there was this very high correlation was quite striking.

Ron Peterson and Cliff Jack have looked at the volume of the hippocampus in relation to Braak staging, and as Mike mentioned earlier, there is a very clear relationship to Braak stage. What was pointed out in this article in 2002 is that, of course, there is alteration in the hippocampal formation in a variety of different neurodegenerative disorders, not just in Alzheimer's disease. So, this type of measure cannot be considered diagnostic, but the fact that there is this very strong relationship suggests that a measure

of the hippocampus should be at least one of the things that you would do if you were trying to using MRI measures as an indication of disease.

And, finally, I just wanted to very, very briefly say something about automated measures of MRI. If we believe that any of these measures are useful, I think it's increasingly clear from the data that I just talked about that probably we need to have more than one measure identified. And, if you're talking about doing this in multiple sites around the country, even if you're just talking about and doing it in a clinical trial, you would want these measures to be as accurate and easy to do as possible. And so, more and more, groups are trying to figure out automated ways of identifying these regions of interest, rather than having to rely on somebody who has a great deal of neuroanatomical skill and with set rules that you could develop and try to train people to emulate over time.

These are data from Mike Weiner's group, showing that they have developed a very nice automated way of identifying the hippocampus. This compares their manual measures with their automatic measures, and these are correlations between the two, which are really quite striking. Correlations of 0.92 and 0.91 are what we would all dream to have.

These are data from Bruce Fischl and my colleagues in Boston, where we have been trying to do a very similar thing. These data were published in *Neuron* about a year ago, looking at a wide variety of regions that Bruce has succeeded in identifying

automatically. And, what we're doing now is to try to extend these automatic

identifications of regions of interest to places critical for Alzheimer's disease, such as the

entorhinal cortex in addition to the hippocampus. This just shows you how far we've

gotten with the hippocampus, so this is a manual identification of the region; this is a

fully automated way. And we have very high correlations between them but, obviously,

we would want to be able to measure multiple regions and not just one.

I just added this one slide as a last slide because of the discussion that we had

earlier about ROC curves, and it seems to me that it's very relevant to this question of the

need to measure multiple regions of interest. These are data that were published about a

year ago by Keith Johnson and Marie Kijewski and Georges El Fakhri, looking at data

from the same subjects that I mentioned earlier. And, what they did was to have to get

nine different MRI regions of interest, superimpose them on SPECT data, and then ask

the question: How do these nine measures in combination discriminate patients prior

to the development of Alzheimer's disease? How well do these nine measures in

combination predict what's going to happen to people over time? And, as you heard

from Keith earlier, the SPECT measures that used to be available were much less good

than the ones that are available now. With improved accuracy, this ROC curve

represents the accuracy of predicting who is going to develop Alzheimer's disease, based

on this combination of nine regions. The blue line represents the data with MRI alone;

the red line represents the combination of the two. And I think it's pretty clear that,

statistically, you can see that there's improvement when you combine the two together.

Even though in fact you're measuring the same regions, you're obviously getting different types of data when you're measuring perfusion versus measuring volume. But, this would never be practically feasible if we didn't have automated measures, so I think it offers the possibility that we might get improved accuracy, but we need a lot more development in technology in order to get there. Thanks very much.

Neil Buckholtz, Ph.D., National Institute on Aging

The next presentation will be by Ron Peterson, who's going to be talking about longitudinal data with respect to MRI in mild cognitive impairment and AD. And then, after Ron's presentation, we'll have an open discussion, again with Mony DeLeon leading the way.

Ron Peterson, M.D., Ph.D., Mayo Clinic

Thanks, Neil. I'm going to now expand upon what Marilyn just had to say with regard to the utility of cross-sectional MR and look at its utility in following people longitudinally. First of all, I need to say that, as you know from doing work in this area, that this a very tedious, time-consuming and expensive operation, because you have to coordinate the clinical groups who are seeing patients longitudinally, hopefully collecting the proper measures, as well as an imaging group that's also imaging the patients longitudinally at about the same time they're getting the clinical evaluations. And, then, if you're going to measure the ability of these techniques to predict disease states down the road, you have to have enough in to be able to generate enough events down the

road to allow you to enter into these prediction models. Most of the work that I'll be discussing will regard the structural MR but, clearly, a whole host of other measures are being looked at, as well, in the world of MR.

And, you know I couldn't give a talk without putting a slide like this. We'll be looking at longitudinal progression of all three groups, and I really want to just highlight the fact that what most of the work done we talk about is done by Cliff Jack in looking at the situation of longitudinal progression of various categories. So, what he's done is he's taken normal subjects, followed them longitudinally over time, and broken them into two groups, those who remain stable, those who progress to MCI or AD. Similarly, the MCI group has been broken into those MCI subjects who remain stable over the longitudinal follow-up period and those who convert to AD. And then, for completeness, have taken Alzheimer subjects and broken them into fast progressors and slow progressors by virtue of did they shift from a CDR category over the time course. Those would be the fast progressors.

And, this is the set of criteria that we use for the memory amnestic type of mild cognitive impairment. This is a scan, again, showing the architecture of a normal subject, pretty plump looking hippocampi, normal ventricular system. Here's a 72-year-old subject with the clinical diagnosis of mild cognitive impairment. Vents are slightly larger, little more generalized atrophy and a bit of hippocampal atrophy but not a whole lot. And, here's a 74-year-old subject with mild to moderate Alzheimer's disease and, again, you can see progression on virtually all of those measures.

The cross-sectional studies that Marilyn has just outlined imply that we're going to see the same kinds of findings over time in the longitudinal studies and that's, in fact, what Cliff has set out to do with this work. In the first paper that looked at MCI progression cross-sectionally, Marilyn has outlined this work already, we looked at 80 subjects with the endpoint being crossover to AD; 27 subjects did cross over, and these are the data that you just saw, again, cross-sectional data. Now, the question comes up if we have an individual subject, will an individual subject from one curve to the next curve to the third curve as the disease progresses? And, this is essentially the same type of question, then, that's asked for the other structures that we're going to be looking at, in addition to the hippocampus itself. And, as Marilyn outlined, a host of other groups, as well, have looked at cross-sectional MR.

The serial studies, though, as I indicated, really are a major effort, and it takes thousands of scans to generate these types of data. And, Cliff has estimated that he's probably done about 28 or 2900 scans over the year, not in different people, many of these in the same people to generate these longitudinal data.

I'm going to talk about four serial studies. The first one just looks at medial temporal lobe in normal controls versus Alzheimer's disease. The second looks at rate of change with clinical status, looking at the hippocampus alone. The third one looks at the hippocampus and three other measures over time. And, then, the fourth one applies this to a clinical trial.

So, the first study was looking at just can we see a change in rates of change now between normal controls and AD subjects. So, we're talking here about two measures. Here is the hippocampus and, then, we're also looking at the temporal horn of the lateral ventricle. And, it turned out that both measures were, in fact, sensitive at detecting a change now. So, this is 24 subjects followed over and this is annualized change, so the hippocampus, the mean change of the hippocampus and the controls and the temporal horn. And, while the temporal horn change looks large; in fact, because of the standard deviation in this particular study, the hippocampus turned out to be the more sensitive of the two measures in following change over one year.

The next study was done looking at change in clinical status and, here, we're getting into the splitting of the different clinical groups. So, the control group again was split into those who remain stable over the two to three, four-year period versus those who crossed over to MCI or AD. The MCI group was divided into those who were stable, those who crossed over, and the AD group into slow and fast progressors. In this particular study, the control groups then, comparing the decliners to the stable subjects, you saw a 2.8% change in the decliners and a 1.7 in the stable group. Similarly, in the MCI group, 3.7 in the decliners, 2.5 in the stable group, and in this particular study, did not split the AD subjects. So, in fact, this is the first indication now that within these clinical subgroups, the changes in volume of the hippocampus over time was, in fact, tracking with the change in the disease state, as if the people were moving down the success of survival curves on the cross-sectional study.

Now, taking the same methodology, expanding the groups and adding additional

measures, so in addition to doing the hippocampus, now looking at the entorhinal cortex,

whole brain volume, and ventricular volume, asking the questions do some of these

measures tell us more about disease progression at various stages. For example, one

could certainly hypothesize that the entorhinal cortex and hippocampus may change

earlier in the disease. So, in converting from normal to mild cognitive impairment, you

might see a change at this stage. Later on in the disease, as it goes from mild cognitive

impairment to AD, you might expect whole brain or ventricular volume measures to be

more sensitive.

These are the tracings of, moving from posterior to anterior of the hippocampus

and the entorhinal cortex. So, these are the structures that, in fact, are being measured

and thought to be involved earlier in the disease, rather than the whole brain measures.

Here again is the temporal horn volume measurement, and here is a boundary shift

integral technique, much like Nick Fox's technique of the red outline here indicates the

change over one year period in volume in a control subject. And, here we see the same

changes in a patient with mild Alzheimer's disease. So the degree of red in this, again,

indicates that that volume has changed over that time. So, here's the control subject, and

here's the subject with very mild Alzheimer's disease.

Busy slide, don't worry about. I'll show it graphically in just a minute. Here are

the four structures being evaluated. Here are the three groups of subjects, the stables

and converters in each column. Basically, all of the measures were pretty accurate at

detecting change in clinical states between those who remain stable and those who converted in each of the three subgroups. So, looking at it in an odds ratio-type fashion, again, we have the Alzheimer's disease subjects, the MCI, normal and the four structures within each group. The odds ratios basically separated those who remain stable from those who converted in almost all of the clinical groups. There was a little bit of overlap here of one but, by and large, they were all pretty good, meaning that counterintuitively, at least to us, in this particular study, the hippocampus and entorhinal cortex changes did not appear earlier or, at least the whole brain and ventricular volume measurements were as sensitive as picking up change as were those medial temporal lobe markers. So, good news/bad news, it's a little bit worrisome in terms of the way we think about the progression of the disease. On the other hand, as Marilyn was indicating, if these whole brain ventricular volume measures, which are automated, can be as useful as measures of the entorhinal cortex and the hippocampus, they in fact may be more useful in implementing large-scale studies.

One implication of this is if one now is going to translate this work into clinical trial design, we could say, well, which measures, in fact, are better at detecting change over time? And, what you can see here is that by putting a couple of clinical measures in here versus the volumetric measurements, that if you're generating power statistics for a sample size to detect a 50% change or a 25% change, taking a look here at the Mini Mental State Exam, these are the sample sizes you'd need in the arms of the study if you're using these clinical measures. If you're using a volumetric measure, say one of

the automated measures, you can get by with a much smaller sample size, implying that, in fact, these volumetric measurements might be better indices of progression of disease over time. At least, they have less bounce in them than do the clinical measures.

So, from this we concluded that MR rates do correlate with the clinical course across the cognitive spectrum, across normal, MCI and AD. But, to our surprise, there was not a great deal of stage specificity with regard to certain structures changing more sensitively at one point in the disease process than others. And, the sample sizes, then, that one would project if one were to do a clinical trial following subjects longitudinally, were much smaller for MR measures than they were for the cognitive measures.

And, there have been a host of other serial studies. Peter Freeborough and Nick Fox, Jeff {Kay} and {Michel Lasso} have done longitudinal studies. A lot of the work by Nick and his group has been on familial AD cases, picking up people who were at risk because of known mutations in the families and following them longitudinally, younger subjects but have they perfected much of this technology.

And then, finally, now applying this to an actual clinical trial, Cliff was involved with a study on memantine a few years ago to study its effectiveness in treating patients with mild to moderate Alzheimer's disease. And, the objective was to see if, in fact, MR could pick up changes in these two treatment groups. Now, those of you who know the study, it was designed as a 52-week controlled trial with an n of 450, but the therapeutic trial itself was terminated halfway through, at an interim measure because efficacy was

146

not projected to be positive. But, the MR part of the study was continued, and 192

subjects completed it, from 38 different centers and basically had two MRs one year

apart, with measures in that study would involve hippocampal and temporal horn volume.

There's a picture of the temporal horn again. And, the study really was a

feasibility study, one that you do it, that you could do a multi-center MR study and that,

in fact, these measures did change over time. But, again, when you pitched these

measures against clinical measures and, then, projected out what would be power

calculations using these data, once again, you came up to the conclusion that to have a

meaningful arm size using clinical measures, you're in the 200, 300+ ballpark, and you

can do them with many fewer subjects per arm if you're using MR measures. So, again,

it was not useful in terms of did it shed light on the efficacy of the study drug, because

the study was halted from that respect. But, it did demonstrate the feasibility of doing a

multi-center trial and, in fact, showed once again that the MR measures appear to be

more sensitive to change over time than do the clinical measures.

So, the technical feasibility was documented. The decline over time was more

consistently seen with imaging, and the power calculations, again, showed what they did

in the earlier study. So, I think these are important in terms of as we plan out future

clinical trials and longitudinal studies.

Marilyn already mentioned this, so I can go through this very quickly. The same

technology now is being appended to MCI clinical trial, and the clinical data, as she has

shown, the very same slides with regard to delayed recall and with regard to paragraph recall. Again, people with the smaller hippocampi recalled less well than those with the larger hippocampi, and the combined groups with regard to progression over time, again, these are the combined groups, but those again with the smaller hippocampi progressed more rapidly than those with the larger hippocampi as it did earlier in that cross-sectional study.

So, I think that MR is useful in cross-sectional studies with regard to supporting your clinical diagnosis. The serial tasks in fact track quite well with the disease progression, so those individuals who are, in fact, going to progress clinically do show that, demonstrate changes on various measures of hippocampus and other MR measures. But, none of the measures appears to be more sensitive than any other, but virtually all of the imaging measures seem to have greater power at detecting differences and changes in clinical state than do the cognitive measures. So, let me stop there.

Neil Buckholtz, Ph.D., National Institute on Aging

Thanks, Ron. I'll ask Mony to come up and initiate the discussion section on MRI. I did forget to ask Marilyn to do any disclosures, so I'll ask her now.

Marilyn Albert, Ph.D., Johns Hopkins School of Medicine

I have no conflict with respect to these specific issues, but I receive funding from Pfizer Pharmaceuticals and GlaxoSmithKline to do imaging studies.

Mony DeLeon, Ed.D, NYU School of Medicine

Since lunch, I haven't acquired any new conflicts, although it feels like I have. I wanted to say just a – actually, this is the most difficult challenge of all, I mean, trying to bring MRI to center attention in a meeting that's designed to appraise the PET. So, I followed the suggestions of Susan Molchan, which I think are very good, which is really to, as the two previous speakers have done, that both Marilyn and Ron have done and also Michael, to show off the highlights of what MR has to offer and where these technologies, which are different, may be converging and where they may be actually having unique value.

This first slide is really meant to represent just the anatomical, spatial and size relationships between entorhinal cortex in yellow and hippocampus in red as seen through this rendering on an MRI scan. These are actually size correct and images. This is where we'd like to be, at least in part, even though this may not very well be the whole story. This is an *ex vivo* study by Joe Poduslo at the Mayo Clinic, highlighting in a gadolinium phase study. This is an MRI study using gadolinium labeled with amyloid beta, highlighting the plaque distributions in the brain. Similarly, this {last} contrasts the study by Einar Sigurdsson and Tom Wiesniewski at NYU, which is done *in vivo* with an intercarotid injection, show plaques but they show much less plaques. And, the third slide, which I'm not even going to show you, is what happens when you administer the gadolinium Abeta 1-40 intravenously. You don't see a lot.

The issue of hippocampus in Alzheimer's disease is one that's been around for ages, since the beginning of Alzheimerology, and it's one that's come to play, as you've just heard, *in vivo* in a number of very important ways, but that's only part of the story. In a slide that Marilyn showed earlier, work that was done with Maciek Bobinski some years ago and in collaboration with the late Henry Wiesniewski, we showed then that there was a very good correlation between the number of neurons and the volume of the hippocampus when either studied *in vivo* or at postmortem. So, we had both scans available on each deceased patient. (Inaudible) the effects were nearly unity; the relationship was nearly unity. As Ron had alluded to and Marilyn as well, the issue is really one of predictions, at least in part. Now, one of the issues is prediction. And, here's an example of a hypothesis that Ron just presented some data on, in other words, which horse comes out of the barn first, and which horse is the most relevant with respect to the diagnostic endpoint which is the crossover of the AD finish line, if you will.

Another hypothesis relates to what happens among normal individuals, which we've also heard a bit about. And, here, you see the hypothesis, although not supported by the Mayo Clinic data, that the entorhinal cortex may be the first horse out of the barn, although there's other data to suggest that that may be the first horse, certainly the PET data that I showed earlier supports that contention. And, there are other data sets that are beginning to emerge, although there's very little published in this arena. So, the question of what constitutes a change in a normal brain that puts a person at risk of future decline is a brand-new field. We must fully acknowledge that, I think.

The MR, because of its inherent capacity to differentiate tissue density, is very useful at boundaries and *ergo* studies of volume and related studies. And, here you see, this is not an MRI. This is a specimen with the ventral surface facing you, showing the boundaries of the entorhinal cortex. And these boundaries can be appreciated on MRI. I'm not going to show you the MRI, but I will show you that in coronal plane, you can identify these boundaries by identifying rhinal and collateral sulci, {sulcal simi} (inaudible). There are a number of features that are identifiable on the MRI scan that come from the newer pathology. In another study that we had looked at for validation, we found a rather good relationship between the distribution, the histological distribution of the entorhinal cortex, which is not per se visualized on a scan, and the landmark based estimation of where the entorhinal cortex actually is. So, basically, you can use a scan to estimate where the tissue is even though you can't directly appreciate the histology.

There's another technique. You heard a few moments ago about Nick's very elegant solution some years ago to dealing with atrophy, at that time at a global level, and it remains a real important mark in the history of understanding how do you use a scan and how to appreciate change in the brain. And, a technique that was derived from that was recently reported by Henry Rusinek at NYU, and let me show you here. This is the hippocampal region, in case, you know, there's any doubt and with a temporal horn as Ron so nicely pointed out the anatomy, done in 1995 and, a couple of years later, you see this same scan from the same person. And, if you actually run the boundary shift, but here, the difference is that this is a regional boundary shift, meaning that you're in

151

essence – it's like going fishing. You take a fishing net, you throw it in the water, and it doesn't matter where the fish is. As long as the fish ends up in the net, you catch the fish. And, so, this is not anatomically specifically driven. In other words, this is not driven by precise anatomy. It's driven by an understanding that fish usually end up in this part of the river and not in that part of the river and, therefore, let me fish over there. And so, with this technique, Henry had gotten some very nice data. What's unique about the data is that these are now, rates have changed by year. This is atrophy rate. And what you see in green are normal people who remain normal and, in yellow, normal people who, after six years, became MCI, and the few that are after six years became Alzheimer's disease.

What was unique about this study, this had an accuracy of 89%, what's unique about this study which is really an extremely important harbinger of where the fields are going and where we're trying to go, is in the observation that within the first two years, while patients were still normal, a subset of patients, I mean, those patients that remained normal for two years and had the scan, baseline and follow-up, showed elevated rates of change, such that by six years, one could predict their MCI. In other words, during the period of normalcy, brain changes are occurring which can be detected. It doesn't require a baseline in the normal state and a follow-up in another state. Within the realm of normalcy, one can make these observations.

Mike had made the point before, and I think very credibly, that he found very little evidence for specificity, and one of the problems in the MRI field, which is even worse than the PET field, is the issue of diagnostic specificity. And, I think this is an area that

absolutely requires our attention. There is considerable confusion. There is some data on

dementia with Lewy bodies, having maybe less temporal lobe atrophy or having changes

in the {cordate} or containment. There's a whole smattering of findings out there that are

not very coherent at the moment. Normal pressure hydrocephalus, well, that's the classic

diagnosis to use in MRI scans to identify big ventricles and compressed cell size, so

that's a good one. You don't need a PET scan there. Frontotemporal dementia, well

there are studies that have shown that they have equal rates of change as Alzheimer's

patients do in the hippocampus and other anatomy, and the differential is very difficult to

do and, with respect to MCI, again, another zero as we saw before with the PET.

So, the issue that I see is really fundamental. It's really this road to specificity.

And there's two definitions for specificity. One is the clinical specificity, the diagnostic

specificity, and the other relates to the anatomical specificity of regional change for MRI.

So, I want to distinguish between the use of that word "specificity." In this case right

now, I'm referring to diagnostic specificity. In work that was done, and I'm showing you

this to – this is really – I feel that I may be off the mark but, you know, it's only going to

take a few more minutes, and here I am anyway – (laughing). This work here, if you

look at the red column, this is at cross-section looking at nine normal and eight MCI

patients studied at baseline and comparing, as John Trojanowski had suggested this

morning, the examination of the imaging with the biological markers that are available

from cerebrospinal fluid. And, so, if you look at the phospho-tau 231, which is the work

of Molecular Geriatrics in Chicago, using an antibody that Peter Davies developed, if you

look at amyloid data 1-40, from {Punkashmet} in Staten Island or the isoprostane F, which is from Domenico Pratico at the University of Pennsylvania, or the hippocampal volume, that red column shows you that you're in the 80s. Your overall accuracy is about 80, and you're doing pretty well.

You ask the question, as Marilyn raised, about how well do these things actually add to each other. Well, the isoprostane measurement boosts the diagnostic accuracy of the hippocampal volume from an already very nice 89% – well, the overall accuracy jumps from 88 to 94%, so it significantly boosts hippocampal appraisal. So, a biomarker can help you out in this game, in these very selective research populations, in these very specific investigations.

The biomarkers also help themselves so, before, Ron was talking about a panel, a panel of imaging markers. Well, a panel of imaging markers is conceptually very similar to a panel of biochemical markers. What do we need to do the job? How much independent variance do they contribute to the diagnostic effort? And, what you can see here is the isoprostane actually contributes to the very nice accuracy of the p-tau (231). So, we get across the bottom row a boost from 82% to 88%, small samples, but do these biomarkers actually talk to each other? Are they measuring the same thing?

Well, it's believed that tau, tauopathy, taupathology in Alzheimer's disease has a very specific start point in pyramidal cells in the limbic anatomy with a preference to entorhinal cortex and hippocampus. We know all of this, but is there evidence?

Correlational evidence sort of gets added, although it's not proof. But, if you see them change in one measure that's correlated with a change in another measure over time, it sort of increases your confidence that you're looking at the right thing, because you can test the specificity, if you will, of that relationship by looking at another measure. The best relationship, the unique relationship that we've found, is that a change over time in phospho-tau 231 actually is correlated over one year with a change in the hippocampal volume. So, here you see MCI patients with a 0.8 correlation of the delta in one measure against the delta in the other. We think this is a way that one biomarker will boost another biomarker by the bootstraps and say, "Hey, you're talking to me, and are you talking to me in these other diseases, as well?", going back to the specificity data at – a diagnostic specificity data that we wish we had. This is one way to examine those questions which I think are not so hard to do.

One of the final points I want to make is that the MRI scan can also help you out in other ways. Proteins end up in the CSF in a variety of manners. They get degraded in the CSF. They have a gradient by which they are often to be measured in the lumbar spine, which is where the LPs are done. And, the process of going from a molecule of tau, from a pyramidal neuron to a measurement of a tau in an (eliza} drawn from the lumbar spine requires considerable assumption about what the fate of these molecules are. So, one question that we had was if you look at the literature of the biomarkers, there's very little evidence for longitudinal change in tau.

So, the argument was, based on some very challenging work that Hansotto Reiber has published in Germany and also later corroborated by {Kai Bleddow} in Sweden, that the levels in the ventricular compartments to tau are two-fold greater than the levels of the CSF lumbar tau. So, ventricular tau concentrations are twice as great, twice as high as are lumbar concentrations. That means then, in other words, that when you're measuring things down in the lumbar region, it's not reflecting the change and what's happening, presumably, and as Ron showed from Cliff Jack's data, you lose some neurons. You lose some volume. The ventricles grow. So, you have this compensation that's virtually for every tau molecule that ends up in the water, you end up with more water, so you end up with up a dilution phenomenon that keeps the concentration constant, a very simplistic explanation, I understand, to a complex problem. But, given that the tau levels are so much higher in the ventricle, we raise the question, well, can we actually correct for this? Can we do something about it by estimating here the ventricular volume? And, by the way, the ventricular volume, as Ron shows and our data show and 50 other people have shown, the effect is somewhere between six and 15 times that of control, and it's growing.

So, if you correct for the growing progressive ventricular involvement, and you end up with a statistic called "p-tau load," you take what was then a non-significant effect, 56% overall accuracy, with a p-tau level seen over one year. It's the delta that's telling you that this chance relationship to group, I'd rather if you just correct for the ventricular volume, you end up with an 89% accuracy. Now, this is not a great solution

to a complex problem, but it is identifying that there is a very important issue, and the issue of CSF flow and clearance and production of these molecules and the fate of these molecules represents a very important arena for research and for the use of imaging to try and disentangle. So, this I see as a frontier issue that relates to this diagnostic question.

A couple of words on diagnostic specificity of the biomarkers. This is work by Harald Hampel and Katharina Buerger from Munich for the p-tau 231 using the same assay I just showed you. You can see here that the Alzheimer on your left shows a rather broad and higher median value than either frontotemporal, vascular, Lewy body, and the control is the very end of the group. So, these data suggest that p-tau 231 is, in fact, elevated in these groups, and in collaboration with that same team, our data show that our MCI patients are also elevated relative to the frontotemporal dementia with respect to the p-tau 231. This is the level, not the load. So, my question really is, and I want to I guess cultivate discussion around is that I think it's really trying to fit these modalities together in trying to develop and improve a way of studying the {American} that's at risk and in need is where do we go? Do we combine MRI and PET, would get something that I alluded to before? Is there some issue? I mean, people obviously don't see it as clear. It's expensive to do this. Do we need, as was pointed out earlier, a CMS trial? In fact, one of the things that I wanted to ask someone from CMS, I understood that there was some demonstration project that was considered at the last meeting. And, this demonstration project, you know, has it been designed? Where are we now in the formulation of that scheme? And I'm very curious about how this larger picture, which

you've now been exposed to today, can be filtered into that prior understanding. So, I

open this to discussion, and I hope I haven't tormented you too badly.

Neil Buckholtz, Ph.D., National Institute on Aging

Thanks, Mony. Well, maybe we'll ask Steve – (laughing) – to respond to that,

first of all, and see where CMS is on it.

Steve Phurrough, M.D., MPH, The Centers for Medicare and Medicaid Services

When we did our original PET decision last year and the Secretary did decide that

we would first have this expert panel and, then, support some kind of demonstration

project without defining what that demonstration project is, and that demonstration

project does have to fall within boundaries of what Medicare has rules and regulations

around its reimbursement scheme. Medicare can't pay for trials. We can't run trials. We

do have policies that allow us to pay for the clinical cost of people who are involved in

trials, using technologies that we provide coverage for. That's a difficulty we now have

with PET for Alzheimer's, since we have a non-coverage decision for that. We have

been investigating our legal authorities to provide limited coverages for technologies

only in the context of a clinical trial and not outside that clinical trial. We would like to

be able to do that. There are some legal impediments to that that we think we can get by.

There are some, obviously some political, both at the national and local level, problems

with that. How do we define who is and isn't at trial? How do we limit people who are

in that trial, or if we don't limit, how do we control that trial? Who would run the trial

158

since our only responsibility – our only ability is to pay for the clinical costs within the trial and not for the research cost. So, there are a whole host of issues that we would like to resolve, because we would like to get to the point of being able to fund the clinical costs of trials. But, all those ancillary issues of running the trial are difficulties that we need to overcome. We don't have legal authority to handle at the moment. We would love to have a decision around a whole host of technologies, not just PET, that says "We're only going to pay for the use of this technology if you're in a trial," because we need more research in a whole host of things. Once we pay for something, research stops. Why do research if you can order it without all the hassle of research?

Neil Buckholtz, Ph.D., National Institute on Aging

Thanks, Steve. So, let's open it, then, up to any general questions regarding MRI. Who would like to go first? Mike.

Michael W. Weiner, M.D., VA Medical Center

I have a specific question to my dear friend, Norman Foster. I know that you're involved in a study now looking at the value of PET scan for the diagnosis of frontotemporal dementia. And, informally, you and I have talked in the past about the potential value of having an add-on where MRI would be done on the same subjects and one could contrast and compare the value of MRI and PET looking at that. I don't know if you've thought more about this or what you think about the utility of such a suggestion?

Norman Foster, M.D., University of Michigan

The study is designed to look just at PET. MRI scans or CT scans will be obtained as part of that study, just as they are following AAN guidelines. We will be able to see, because all of these images will be digitized, to what extent there is a correlation between the regional MRI atrophy observed by our consensus panel and the abnormalities that are seen on PET. So, I hope that will provide some additional information.

Male Voice

It's a question for anyone who wishes to answer, CMS perhaps. An argument that I've heard from some PET (inaudible) such as academicians and all of PET is that why is the bar different for PET than MR? MR is primarily used as a diagnosis of exclusion, to exclude other disorders. It was very, very quickly approved for reimbursement. But, when PET got to the station, the bar went up and is being asked to demonstrate – I'm not taking a position on this, but this is sort of an argument that has been extended. It's a diagnosis of inclusion and, in some ways, sort of facilitates. You can argue about how much incremental value, but it's become a debate that's never-ending, it seems, going on year after year, different sort of groups, different set of researchers convened, policy statements, etc., and not sure where it's going, so if someone could comment on that.

Steve Phurrough, M.D., MPH, The Centers for Medicare and Medicaid Services

Because we made reimbursement decisions inappropriately in the past does not

necessarily mean we should continue that – (laughing). We've previously paid for a

whole host of things that we shouldn't pay for. It came on the market, and we paid for it

and made very little review as to whether we should or shouldn't. We are now applying

a much stricter evaluation process to people who want us to pay for that technology. I

think that's appropriate. I think that's the way we ought to go. You certainly can argue

that we may be making incorrect decisions, but I think it's appropriate for us to at least be

going through that decision-making process.

(End of Tape 3, Side 1)

Neil Buckholtz, Ph.D., National Institute on Aging

Other questions regarding MRI? Going once –

Male Voice

Also, I just wanted to indicate that I believe with the use of MRI, just as with PET,

the important thing is to take clinically relevant questions to try to answer. And, one of

the problems with many of the studies, both in MR and PET, is that they have been trying

to answer questions that may not be so clinically relevant, or at least the studies have not

been designed to answer questions about how to use them in practice. So, even though it

hasn't been mentioned, I wanted to point out that changes in brain volume can be seen in

many other things other than neurodegenerative disease, such as dehydration and aging.

Aging is a major effect and, so, if we were looking at asking Claudia's question about the

over 90s, looking for hippocampal atrophy would be very common in that group. And, many of the data that we saw with regard to MI had to do with normal versus AD as one kind of dementia. And, I think it's important to see to what extent diagnostic tests can improve the accuracy and competence of diagnoses, and there's a lot of work still needs to be done, as you mentioned, Mony, with MR in this regard.

Mony DeLeon, Ed.D, NYU School of Medicine

Neil, a question that I have, following up on Norman's observation, the NIA/NIH has a portfolio of longitudinal MRI studies in this arena. I don't know how large it is, but I'm sure there are several, from what we heard today and others that are not here. And, there's probably a portfolio, as well, but a smaller one for FDG-PET, making it very burdensome for the NIH to say, okay, we have this very interesting longitudinal MRI data, and I'm presuming now, on a larger set of individuals. What would it take, for example, to link the ongoing NIH studies, NIH-supported studies, with a CMS initiative that would enable, at least at the beginning, for those studies to add on a PET scan or even a longitudinal PET scan so as to improve this issue of sensitivity for earlier detection, the course of the illness and, ultimately, the differential diagnosis. Is that kind of cooperation between the Institutes, is there any precedent for cooperation between CMS and another Institute, because the R&D has largely been accomplished. If you look at those studies, there's a tremendous amount of technological development that's already in place. So, as a starting point, it's a low-cost solution to a very complicated problem.

Neil Buckholtz, Ph.D., National Institute on Aging

It's a very interesting idea. I don't know if Steve wants to talk about whether there are any precedents that you know of?

Steve Phurrough, M.D., MPH, The Centers for Medicare and Medicaid Services

We've funded a lot of studies. We specifically funded the clinical cost in a lung volume reduction surgery trial that NHLBI sponsored. We worked with them to develop that trial, five-year trial, changed our coverage criteria to allow coverage within that particular trial. And there are other trials where we don't necessarily have to cooperate with NIH, we just provide the coverage of those clinical costs. We certainly are interested in (inaudible). The issue of a demonstration, just to define a "demo" that will relate to my next comment, Congress has given Medicare the authority to do what it terms "demonstrations," but that it is a very narrow type of evaluation. It's essentially to decide whether you can do something more cheaply, using one kind of care provision versus another kind of care provision. So, it's around disease management and those kinds of things. It really doesn't fit what we would like to provide reimbursement for. So even though the Secretary said we do a demo, the kind of answers we'd get from a demo are not very helpful. And, so, we are interested in cooperating with someone who, in fact, can design a trial, and like we can't, that would allow us to get the information we need. And, if NIH has some of those trials that could be modified, that would add PET in

163

a manner that would answer some of our questions, then, yeah, we'd certainly be interested in seeing if that would work.

My question and perhaps one that we could get some comment on is what is the ideal trial? What kind of design should we use, based upon some of the R&D that you've mentioned has already been done, kinds of pieces of data that we should be obtaining from the patients, both clinically and ancillary wise, markers, imaging and so forth. How long should it last? How large it should it be? And those kinds of questions are questions that we're certainly interested in to make sure that if we're going to fund something, that the right questions are answered.

Mony DeLeon, Ed.D, NYU School of Medicine

In my judgment, those are answers that are not at the table at this moment, but they could very readily be collated by sampling the people in this room and others and convening perhaps one or two more workshop-type environments and, then, deriving a best estimate of the kind of study that could go forward and tackle that very type of question you wanted to raise. I'm here for a moment. I'll be gone in a second. And so, in my moment, I'm just curious about the view that Dr. Hodes has – (laughing) – about this – you know, I see you in the back there, far away from the microphone, and I wonder about the issue of cross-communication between CMS and the NIA. Is that a vista that we should really work towards developing in this arena? Do we actually, you know, write memos to ourselves and to our political leaders and get this ball rolling?

Richard Hodes, M.D., NIA

I think the answer is certainly that we wish to, should, and will pursue that. Actually, I'm not sure it's at a stage where writing letters is effective, and I don't know that I'd hesitate to say that it was if I thought it really was. But, I think what a number of people have asked here is most appropriately what is the study, the research question that needs to be answered, how should that be designed. And, once we're at that point, I think we can ask what the role would be for CMS cooperation and coverage. But, I think the question that CMS is asking us – what is the optimal study– really it has to be what comes first. I don't know that, as you've heard, CMS can carry out demonstration projects, not research. They therefore would not be actually involved in the research design. We would work closely with them, because the end product of research would influence their practice. There might be a role, as was alluded to, in their assisting financially in the carrying out of the study, but I think the protocol designs, the posing of the scientific questions do need to come from the people in this room through research agencies, and NIA is happy to be playing a leading role in that regard.

We're also I think very cognizant of the question you've been asking about collaborations, metanalysis, add-on to longitudinal studies already under way, as well as those that, as you know, are in the process of being reviewed for funding, to make sure that the data that we are generating are compatible or cross-comparable and will give it the most powerful answers to the relevant questions being asked. So, I think, yes, this meeting to be working with CMS is very relevant. I'd be happy to hear a constructive

suggestion for how letter-writing might help, but I don't know we're really at a point where we need new authorities other than to focus on, perhaps, sharing in financial support of some of the standard of care. Otherwise, this is the group. NIA is happy to be playing a role that should be designing what the optimal study is and then getting it carried out.

Neil Buckholtz, Ph.D., National Institute on Aging

Thanks, Richard. Yeah, I mean, I think this is the beginning of some other continuing discussions in terms of looking at the feasibility of this kind of study and, as you say, whether there are ongoing studies that we could add to or whether we need to develop an entirely new study, whether the new imaging initiative may contribute some to this. There are a lot of questions that we really have to follow up on, based upon the information that has come out today.

Richard Hodes, M.D., NIA

I actually misstated it. I stated it as longitudinal MRI that could have PET added, but it should certainly go the other way with PET studies have MR added, just to go both ways with respect to the new side of collecting this information. Sorry for the interruption.

Male Voice

(Inaudible/off microphone) to maximize the earlier studies to (inaudible) talk to one another (inaudible) capabilities or at least harmonization in their context. And, some of you have probably been at the receiving ends of these kinds of inquiries as studies have been proposed. So, the point is very well taken.

Neil Buckholtz, Ph.D., National Institute on Aging

Yes, I mean, more and more within the Institute as we are funding a variety of different kinds of studies, the question arises in terms of issues of standardization, of methodology, and especially issues of sharing and compatibility of data. And, so more and more, as Dr. Hodes said, you're on the receiving end of questions from us in terms of ability to share data, to make your data available to other investigators, to be able to look at these kinds of issues. So, I think that' becoming more and more important.

Male Voice

To throw out an idea that could easily be shot down, if I was trying to leverage existing resources, what if we used our national resource of ADCs, individuals are followed longitudinally to autopsy, and some subjects who are referred to primary care physicians and dementia specialists outside the program who have no access to PET and MRI? Some have access to PET and MRI and see how that affects the judgments of the physicians and predicts course and differentials.

Neil Buckholtz, Ph.D., National Institute on Aging

So that's something our National Alzheimer's Coordinating Center, which coordinates the data among all the centers, might have those kinds of data available, at least to take a look at that.

Male Voice

It would seem, with respect to the two questions that are in front of CMS right now with regard to PET, one is differential diagnosis, say Alzheimer's disease versus FTD. And, you know, Norm has a grant out there, an RO1, that's looking at that issue. There is another RO1 with FTD being studied among about five of the Alzheimer's centers looking at instrument development for FTD, which seems like that would be a nice avenue on which to tack on scans if it were feasible, if the ends were large enough so you'd be collecting a couple sets of data. And, with respect to the other question that CMS is addressing, that is, early detection of AD, so MCI, it would seem that the Neuroimaging Initiative, while focusing on MR, certainly could be tweaked, since the mechanism should be there to recruit the subjects, which is the main cost effort, that one could tack on PET scanning – more PET; there's some being planned in the various proposals under review. But, I would think with some enhancement, perhaps of a funding from CMS, that might be more feasible. So, there are mechanisms out there of ongoing studies to address the two questions that are being put to us.

Neil Buckholtz, Ph.D., National Institute on Aging

Just to remind people, in the Neuroimaging Institute, we will be looking at both MRI and PET. Everyone will have MRI; a subset will have PET, as well. And, so, we're not sure exactly what percent that's going to be, but I think a reasonable percent also, they'll have serial scans in both.

Female Voice

Just to repeat what we talked about this morning, the Neuroimaging Initiative is going to have very carefully diagnosed subjects, and the real question that's been coming up all day is what would happen in a primary care setting, because that would really be the implication of any approval by CMS.

William H. Theis, Ph.D., National Alzheimer's Association

Well, I might just make two comments. I'm Bill Theis. I run the science operation for the Alzheimer's Association. And, as someone who often gives staff instructions like, "Do something good," I've learned that a little more specificity helps. So, in terms of the letter writing, it doesn't do much good to write letters until we actually know what we might be writing them for. And, the second thing is in terms of the trial design, I found some of the presentations given by Dr. Matchar to be fairly compelling. With a model of that sort, whatever the initiative is is going to have to inform or modify that model, I think, to give different results, because that model clearly showed that with some improvement in the diagnostic ability, that was not transferred to the endpoint of the patient outcome. So, I think a lot of the existing imaging that's going on is, at least to

my knowledge, not quite so strongly oriented toward patient outcome. And, I think if we don't keep that in mind ahead of time, that's going to be a deficiency.

Neil Buckholtz, Ph.D., National Institute on Aging

One more question, then we'll go on.

Male Voice

I hate to be the stick in the mud in this all day long, but there are tremendous market forces that are going on right now throughout the country in terms of the impact of diagnostic imaging on hospital bottom lines, on hospitals moving from, say, a 1.5-tesla machine up to a 3.0-tesla machine. And, they're a secondary, community care hospital. PET imaging has become a very important model for cardiac and cancer care, and there are forces in the hospital boards looking at how, if this were to come forward in terms of another indication for reimbursement for PET imaging, how this would favorably impact the hospital industry. So that Dr. Matchar's series of questions that he posed to us this morning really become, I think, the next stage of services research that needs to accomplish before we can reasonably inform CMS in terms of whether or not this is a good deal to improve patient care for a cohort of folks that are moving down the pipeline over the next 25 or 30 years.

Neil Buckholtz, Ph.D., National Institute on Aging

Thank you. I think we'll go on now to Claudia Kawas' presentation on treatment and implications, including potential disease-modifying treatments. Let me just point out that we don't have a formal break this afternoon, but there will be refreshments out there about 2:30, so people can go in and out as they wish for the rest of the afternoon. Talk about the things that are critical, you know – (laughing)?

Claudia Kawas, M.D., University of California Irvine

While we're getting set up here, I have no disclosures to make; neither pharmaceutical firms nor imaging companies have been interested in sending any money my way. In fact, the only money that's ever been sent my way for my entire career is the National Institute on Aging, who I am grateful has funded all of the research I've done. And, I'm also grateful for their inviting me to this meeting, because it's actually turned out to be terrifically interesting and informative. I think everyone agrees.

I am going to start out by saying that nobody – nobody doesn't think that correct diagnosis isn't very, very important, and it doesn't matter what disorder we're talking about. For me, "correct diagnosis" means a diagnosis that ideally predicts the clinical course, ideally identifies the pathology, ideally predicts the appropriate treatment for a particular patient. And, when I started in this field, actually it's almost a quarter of a century now that I've been doing nothing but dementia work, I really thought that I was not going to have a job as a clinician within the next few years. I mean, at that time, Peter Davies and others were coming up with Alz-50s and all sorts of markers, and there

was just no question; I think most of us thought that we were going to have a blood test, an X-ray or something that would make it so that I wouldn't have to spend three hours in the clinic with a patient trying to figure out what their diagnosis was. And, a quarter of a century later, I'm still doing those three hours.

And, since I'm not a very patient person, I would be the first to say I can't wait until we have a biological marker or a set of markers or imaging that will make it easier and more accurate to diagnose people. I have patients right now in my clinic that I would love to see PET scans on. But, a lot of what motivates me is more interest actually, I think, than any real data that affects the therapy and the implications for the particular patients. But, that's what I was asked to talk about was the {fairer} treatment and its implications, and I specifically was given the two questions here by Susan. One is: Does early diagnosis with PET or anything, for that matter, make a difference in treatment and patient management? And, a second question is: Does the disease-modifying drug change the situation? So, being very data driven, I went and I looked to find data, and I'm going to apologize in advance. I found very little. I'm going to show you what I found. And, maybe if anyone in the audience knows of other data that's available that speaks to this issue, I'd love to hear about it.

But, this was one of the first papers that I found. It was published just a little over a year ago, and the title sounds like it's going to give us the answer to the question that I was supposed to ask, "Long-Term Effects of {Rivastigmine}," but the key here is that this is analysis that was done in moderately severe patients. All the patients in this

analysis are global deterioration scales of five and six, so we're talking about very severe dementia. Also, problematic with this analysis, as well as the others I'm going to show you because they are all the same, is about 10 words in the first red bullet point that make a person like me cringe. And, these are post-talk analyses; let's start with that. They were superimposed on the parallel group with Rivastigmine study, which was a 26-week placebo-controlled study, which is very good, but it's only the completers. So, we're only talking about a subset of the individuals in the study. It's an observed case analysis, and it's done during open label. So, I mean, almost any single one of those things makes me cringe, but having said that, they wanted to look at the long-term efficacy, stratified by baseline dementia severity, and they wanted to look at the benefits of early initiation, and this is sort of an oxymoron, early initiation but in moderately severe people – (laughing).

And, this is the gist of their design. I apologize. I told Susan that my definition of an expert is somebody who's willing to talk about something they know nothing about, and I have reached that exalted level by giving this talk. I put these slides together just about 24 hours ago so left out a few important details.

This is actually the design of the study. All together, we are talking about 158 people out of an almost 1,000 person data set that was combined from international and national in the U.S. studies. And, here's the basic design, so if these are the severe patients, the 150-some-odd patients, some of them were on a high dose. Some of them were on a low dose of Rivastigmine, and some of them were on a placebo. And, here is

where they ended up at the end of their 26 weeks. At that point, all of them are given an opportunity to go on open label, and I guess now that the Rivastigmine patients know they're on the Rivastigmine, they sort of seem to, if anything, they level off. Once again, the dose, by the way, to be changed in here, and people were receiving between two and 12 milligrams. You see the low doses of Rivastigmine patients stay the same, and the people who were on placebo without question moved back towards the patients who were on drug all along, but they do not quite reach that level. And, that continues out for a full year.

And, from this data set, the study reached the following conclusions. One was that patients with moderately severe disease had the same benefit for one year, and patients treated from Day 1 with moderately severe disease, though, had greater benefit compared to those who received only six months of treatment. I'm not quite sure how to interpret that really. Maybe that's more of a longevity affect, that is, people who were treated for a year might overall look better than people who were started out at the same level and only received six months of treatment.

On the other hand, we assume this is symptomatic treatment, and one wouldn't expect that by duration but, anyway, those are the results. And, the treatment effects were more robust in this subgroup, that is, the CDR 5, the severely demented people, than it was in the rest of the patients who were mild to moderate in the trials. Maybe this means we could wait until people have moderate and severe benefit, because they'll get more results. I don't know. But, in fact, I want to point out that that's a finding that's

been very, very commonly reported in all of the cholinesterase trials that have been post-hoc reviewed, which is that even though in the original {Paquin} trial, even, when we said you couldn't come in if you had a Mini Mental less than 10, as we do in most studies, we did so because we thought, with a cholinesterase inhibitor, you probably wouldn't {have a respond} if you were so severe in your disease. And, yet, the individuals down at the lower Mini Mental ranges always show a larger treatment effect, at least on the instruments we are using. I have never been clear on whether that means that, in fact, people with more severe disease do get better more benefit, which is possible, or if it's just a product of the measures we use and their non-linearity and how it's much easier to get a four-point improvement on a Mini Mental of 10 that it is to get a four-point improvement on the Mini Mental of 27, right? So, some of this might have to do with the statistical properties of our studies and our designs, rather than a real effect, but nonetheless, overall cholinesterase inhibitors have been shown to show larger effect sizes in more severely impaired patients actually.

This is another from the same study, another post-hoc analysis. This time, we're going to be looking at all of the subjects who were mild and moderate, as well as severe, and came into the trial. Once again, it's the completers. Once again, it's an observed case analysis rather than a last observation carried forward. In this study, however, in this post-hoc analysis, they stratified for hypertension. About 25% of the subjects in the trial had hypertension. That right there tells me, without looking at anything else, that they're a relatively selective group. By age 75, 50% of the population has hypertension.

So, we're either in a much younger age or else we're in patients who've been screened out with a lot of other diseases, in all likelihood. On that hypertensive stratification, however, they did find a significant treatment difference for the hypertensive patients on two scales, the DVS and the progressive deterioration scale. There was a little bit of a signal trending in the right direction but not a significant result, however, for the ADAS-Cog, their cognitive measure. They found no difference, however, in the majority of patients, three-quarters of which were not hypertensive, and there was no difference between being an early or a late starter, if you were non-hypertensive. That is, after six months, if you were given the drug, you went up to the level of people who had been taking the drug for 12 months.

Their conclusions, then, were in the non-hypertensive subgroups, outcome scores in the original placebo group tended to catch up with those in the early starter group. But late starters in the hypertensive subgroups failed to catch up with early starters, and they postulated a vascular neuroprotective effect. This is a subgroup – the people that we are talking about that generated this result are approximately 40 patients in each group out of thousands, literally. So, we really parsed up a lot of things here, and you have to be a little careful when you draw conclusions like that. We all know about studies like the ALCAR study, which showed a dramatic, or so we thought, effect in young patients, but then, when we – the young subgroup, but then when we try to replicate it, we often find these post-hoc subgroups are a poor way to be sure of the results.

Now, I'm going to switch – I'm starting out, obviously, with the treatments that are available, because that's, as a clinician, all I have. At least from a pharmacologic standpoint, the only other treatment we have available is an MDA, memantine, and to the best of the knowledge, there is no data available that suggests that the early treatment with that has significant benefit for patients. It's an interesting area, however. I mean, to date, the only data I'm aware of is unpublished in mild patients. It was an add-on study, where memantine was added to cholinesterase inhibitors. And, in a fairly short duration study, they did not show any benefit with memantine in mild patients, which has a lot to do with why the indication they received from the FDA was for moderate or severe patients. And MDA, as you all know, however, receptors have the potential to maybe hurt or help the brain, and there is a possibility and it has suggested that memantine may be neuroprotective. If indeed it were, then one could argue almost any neuroprotective agent you'd ideally like to start sooner rather than later, but there's no data available on that.

The third and more important form of therapy, though, that I think I give when I see patients in my clinic is not necessarily pharmaceutical. And, most of the time that we spend with patients in the clinic and most of the things we do for them have more to do with an additional benefit that I actually think they get more out of than they get out of the cholinesterases and other drugs that I prescribe.

So, are there benefits for early diagnosis to the patients and to the caregivers? Absolutely. I mean, everybody knows the story of how anxiety-provoking it is for a

woman who gets a mammogram and is waiting for the results of a breast biopsy, or a person who has a CAT scan that shows something that then needs further work-up. Families and patients have a lot of stress over not knowing the answer, in many cases and, certainly, being able to tell them the diagnosis in many cases will relieve stress and will facilitate their own personal planning. Having said that, I think that – Pat Lynch is sitting next to me and pointed out to me that this is what we've been referring to throughout the day as the value of knowing. But, not everybody has as much value of knowing probably as the data-driven audience that's sitting in this room today. And, I think a good demonstration of this that really brought it home to me was a number of years ago, I was at Hopkins when the first genetic testing for presymptomatic or asymptomatic patients with Huntington's in their family became available at Mass General and at Hopkins. And, the most stunning thing about that to me was much to my shock, the overwhelming majority of people who were eligible to have this testing opted not to do it. This was a real surprise to me. We're not talking about Alzheimer's disease in 90-year-olds. We're talking about young people who have a 50-50 chance of getting a really horrible disease at a time that might be relevant for their childbearing and other issues. And I had thought that a fair number of people, as brutal as it was going to be from a counseling perspective, that a fair number of people would want to know. And, in fact, we could tell them unequivocally not only whether or not they'd get it, but we could tell them probably more or less when they'd get it. And, only about 30% of people to whom we offered this genetic testing to opted to take it. So, just because we want to

know and just because we think that the patient might want to know doesn't necessarily mean that those values are shared by the very people we're trying to take care of.

Nonetheless, for me personally, I find it very difficult to tell somebody "I don't really know if your memory problem is going to be a serious issue. I don't really know if you're going to be worse, a lot worse a year from now or the same or maybe even better. And, I wish I did, but we'll just have to take it day by day, and I'll watch you." The only thing I can think of telling someone that would be worse is to tell them that they do or are going to have Alzheimer's disease, and then they don't, or vice versa. So, in terms of giving diagnosis, I think it is very important when we're talking to individual people, the group data, all bets are off. All that matters to that person that's sitting in front of us is whether or not we're right about them. And, that person who's sitting in front of us will have more disservice done to them by getting the wrong answer than they would have by getting the "I don't know" answer, I think in all likelihood. I think many of you and many people in this room were, in fact, quoted in a *Newsday,* or something like that, article about exactly this issue. I mean, where one person was told he's got Alzheimer's disease and, 10 years later, he's now rejoining the world because he hasn't developed it. And another woman was told, with the same scan, two different interpretations. So, I think that kind of problem is actually very, very crucial for us to keep in mind as we're choosing instruments that we will call diagnostic tools rather than research tools.

So, does early diagnosis make a difference in patient management? Well, I guess my answers are probably not a major difference for the available pharmaceutical

therapies right now, and probably a significant difference, a major one definitely, for non-pharmaceutical management, for deciding whether or not to report somebody to the DMV and have their driver's license taken away or the Power of Attorney or whatever.

Now, would a disease-modifying drug change the situation? Absolutely. And, I'm going to make the suggestion that even though we all agree it would change the situation, how it would change the situation is going to be completely dependent on the specific treatment and, in particular, timing. For a long time, I've thought that timing is a lot, and we kind of act like, well, gee, if estrogen does something at this time or it doesn't do something, that it's going to do something at other times, too, and I think that the human body is a little more complicated than that. And, it marches age for a reason, and I think that there's a lot of timing issues that we need to pay attention to.

I'm going to show you right now just a little bit of data from Frank LaFerla's lab. Actually, this is not published yet. And, there's nothing really special about this data, except that it's one example of how I think timing is going to be important. So, these are transgenic mice. These mice are triple transgenes, in fact, so they're both APP, presenile and n-tau {mouse}. And, because of that, they develop plaques and tangles, which is very nice model. And, as you all know, there's an awful lot of work going on, taking all kinds of transgenic mice and giving them Abeta immunization. And, when you do that, we all know that the amyloid disappears, at least for a while, and that's a very promising area of therapeutic work.

It turns out if you give tau and try and inoculate these animals to tau, they do not get rid of their tau, and that's shown here that it doesn't affect either Abeta or tau when you do an anti-tau injection. But, in these mice, it turns out if you give them an Abeta injection, it turns out that you not only affect amyloid but, in these mice, you can see right here, the neurofibrillary tangles, here's a contralateral control side. You can see the neurofibrillary tangles almost completely disappeared in this transgenic mouse that was inoculated with an Abeta injection. So, Abeta inoculation cleared not only the amyloid, but it also cleared the tau, which is – Bob Terry is going to strike me dead when he hears me say this, but this, to me, is the most impressive amyloid cascade hypothesis evidence I've ever seen.

Now, it also turns out that the clearance and ultimately the reemergence is hierarchical, so that first the amyloid disappears and then the tau disappears. And, here you can see, this is three days post-injection, and three days post-injection, the amyloid has disappeared, but the tau is still there. The tangles are still in the layer. However, 30 days later, the amyloid starts reappearing, okay, but the tau, all the tangles, you can see, it's a nice whiteout there, has not reemerged yet. So, tau reemerges at 45 days, but timing is everything. And, after a series of additional experiments, in particular, Frank is ready to make the statement that it doesn't look like the tau burden is cleared, unless you catch it before it is hyperphosphorylated.

So, that brings me back to the other question. Well, my opinion, to summarize, is that there's no clear data to suggest earlier treatment with cholinesterase or memantine in

mild patients necessarily has any efficacy or significant benefit. We don't know, however, if there's any neuroprotective effects long-term, especially with memantine. Misdiagnosis can be harmful in any direction. Future therapies may change the balance, particularly if those therapies are disease modifying, but I don't think disease modification will occur equally at all time periods. And, because of that, earlier detection will have utility only if the detection of the disease occurs at the stage when the particular intervention that we want to put into place is effective. So, just pushing our diagnosis early and earlier, which I can't help but note we've done very effectively, Bob Katzman once told me that, about 15 years ago now, in the California state Alzheimer's system, the average Mini Mental of patients being diagnosed was about 22. Now, in that same database, in the same centers, the average Mini Mental where people are being diagnosed is over 26. So, we've detected people earlier and earlier, and we've improved our accuracy during that time period really quite remarkably. But, until we detect at the place where whatever intervention we have actually works, then early detection won't accomplish all that we would like it to. So, that's it.

Neil Buckholtz, Ph.D., National Institute on Aging

Thanks, Claudia. Again, we're going to hold off questions until the end of this presentation section. The next presenter is Scott Gazelle from the Institute for Technology Assessment in Boston. Actually, Scott, maybe we should take about a two-minute break – (laughing). I want people to be –

G. Scott Gazelle, M.D., Ph.D., National Institute on Aging

It'll probably take me two minutes to – (inaudible/mixed voices).

Neil Buckholtz, Ph.D., National Institute on Aging

– in the room when you're making your – so, let's – all right, let's take a two-minute break and, then, just get something and come right back if you will.

(Break)

(Rest of Tape 3, Side 2 blank)

Neil Buckholtz, Ph.D., National Institute on Aging

– will also make a short presentation with respect to perspectives from the Alzheimer's Association.

Bill Thies, Ph.D., National Alzheimer's Association

Well, I'd like to thank Neil and Susan and Steve for giving me a few minutes on the podium, and I'd like to let you know that I personally am very comforted that Neil and Susan wore their uniforms today – (laughing) – as a visible commitment to sacrificing themselves first if hostilities break out. Then, in terms of conflicts, I don't have any, unfortunately. However, if there's anyone in the audience who would like to test my ethical resolve, please see me afterwards.

Basically, I'd like to tell you a little bit about what I think is the official statement of the Association, and it's official until my boss says it isn't. But, the fact is that it represents the conclusions of a large group of people, and the great part about my job is I get to talk to really all of the world's experts and most of them will speak to us. So, I think that this is a consensus of opinions across the field, and the consensus is perfectly clear. It is, one, most people with Alzheimer's disease don't need a PET scan and, two, some people, a few people would be benefited. Now, the problem has come up, because the CMS guys have said, "Well, that's all well and good, but how do you tell them apart?" And, of course, I wish the dialogue had stopped after we got the consensus, because it gets a whole lot more complicated after we get to "some and a few."

Now, basically, the Association thinks that it's okay to reimburse for a PET scan after there's a complete work-up and the diagnosis remains uncertain. This is basically the essence of what Mike has already presented to you in some detail from the {work} group. I think, on the other side of this, we remain somewhat concerned that broad reimbursement can lead to some inappropriate use. And, I have to tell you, we went around and around about inappropriate use, what the right word was; it started out as "abuse" – (laughing). However, I think that "inappropriate use" is okay.

The success of any kind of reimbursement indication I think is not going to take rocket science to evaluate. If, in fact, the result is a small number of patients with clearly unusual and puzzling presentations are being scanned, I think that's probably about right.

On the other hand, if large numbers of clearly demented 85-year-olds are being scanned, that's probably not a good use of this technique.

Now, in terms of consequences, the inappropriate use is not without consequences. It includes some unnecessary exposure to radioactivity, clearly unnecessary use of medical resources, and what I think is a real risk in the general practitioner pool is some confusion about what's an appropriate diagnostic process. Should you just send everybody for a PET scan when you order their laboratory tests?

In addition, I think there are some further risks here. You all know there are other parts of the imaging community where there's a fair amount of marketing that goes on to the worried well. There's really no good medical indication for testing those people but, in fact, if you drive around in your car and listen to the radio, someone will come and tell you you need to have some part of your body scanned or you may not make it to the end of your trip.

Exploitation of primary care physicians, marketing that implies that this is the way to get out from under what is sometimes an uncomfortable diagnosis for primary care folks or a diagnosis that they may be insecure about making. And, finally, sort of the marketer's dream here is where we end up with a PET scanner in every mall. I don't think anybody wants that to happen and, clearly, that would impact many of us. There probably isn't anybody in this room who isn't going to get the finger pointed at them if that's the outcome. It's going to include industry getting labeled as greedy, government

agencies getting labeled as not paying attention to expenses, nonprofit health agencies like us looking like boobs that are easily manipulable. The enthusiasts have the risk of backlash, and the skeptics have to wait years to get their true insight appreciated. So, this is not a good thing to happen, and I think most of us would agree to that.

In terms of trying to guard ourselves against these positive/negative outcomes – positive/negative outcomes, that was pretty good. I've never used that in a sentence before, these negative outcomes. I think that there needs to be continued study, and the sorts of questions that come to my mind are it's clear we have some idea of how this contributes to a diagnosis, but I think we need to know more. We need to know more about the appropriate patient selection. We need to know about what patients are going to be benefited by this test in terms of treatment selection. It'd be nice to know a little bit about patient satisfaction, although that's not the driving force. And, the fact is that I think if we continue to observe this test in a controlled fashion, we will find some other issues that we haven't even thought of as being important yet. And, so, we need to have the sort of setting that's collecting enough data that allows us to do this.

So, to end all this, I'm going to come right back to where we started. We think it's okay to reimburse in a very closely controlled setting. We are concerned about inappropriate and overuse, and we're committed to continuing in this dialogue to try to come up with the best possible set of indicators and to continue to move PET and other early diagnostic techniques into the forefront of Alzheimer's disease. The Association's going to continue to fund research. We were early funders in most of the FDG-PET stuff

and are happy to be there. But, in fact, there is a different set of questions necessary for the inclusion of these techniques into routine medical care, and we're committed to being a part of that dialogue. Thank you.

Neil Buckholtz, Ph.D., National Institute on Aging

Thanks, Bill. The final presentation will be from Anand Kumar, giving the perspective from the American Association of Geriatric Psychiatry.

Anand Kumar, M.D., Neuropsychiatric Institute, UCLA and American Association of Geriatric Psychiatry

Well, thank you for staying. Everything has been said, but not everyone has said it – (laughing) – so I shall spend the next few minutes going over some key points, sort of the punch line being the perspective, you know, our organization has not, sort of does not have a policy statement, a formal policy statement on this like the Alzheimer's Association. But, the perspective we bring is one very closely related to the Alzheimer's Association, sort of where this should go, where it is now will be very comparable in our recommendation.

I'm here today representing the American Association of Geriatric Psychiatry. My day job sort of I'm involved in studying late-life depression and using neuroimaging, mostly MRI measures, trying to understand the biological basis. The only – it's not a conflict really, what you should know is I work at the UCLA in the same division of geriatric psychiatry that Gary Small does, and Gary and I have chatted about this –

(laughing) – on occasion. At the end of my talk, you'll see how close or how far my position is from Gary.

The AAGP has about 1750-1800 members. About 1500 of the members are MB psychiatrists. The other 200 and 250 are psychologists, sociologists, etc. Almost everyone is closely involved in patient care. While there are many practitioners and clinicians in the membership, the Board of Directors discussed this issue a few weeks ago for the second time in the last couple of years. It is well represented with academics, some of whom, (inaudible) {Colenda} are deans of medical schools, researchers and so on. My own funding is from NIMH, though I would be glad to accept to accept funding from the NIA at any time.

We were specifically asked to keep away from sort of scientific kinds of considerations, comparison of the data and so on and provide some broad policy statements. It's hard to talk about policy statements without what I think talking about the elephant in the room. And, some people have sort of grappled with it to different degrees in their presentations earlier today. The healthcare environment is something most of our members, most of the practitioners involved are very concerned about, as is everyone in the room and as is everyone in this country. This is a title, "Finite Resources, Infinite Need," so I actually plagiarized that. It's the title of a book written by Bill Kissick, who's sort of a healthcare economist at the Wharton School who wrote this book about a decade ago, and he talks about the difficulties Medicare faces. This one here is about every day. In addition to that, this is what practitioners, department chairs

and medical school deans, etc., sort of struggle with, rising healthcare costs. And, high-tech imaging has been identified by several groups as sort of something that's gone up dramatically in the last few years. *The Washington Post*, *New York Times*, etc. have sort of written several pieces indicating that while Medicare has had some impact – reimbursement has some impact on physician services and in reducing length of hospital stay, they argue that high-tech imaging has gone up dramatically. This is not restricted to PET scans, not restricted to the central nervous system, but across the board. It's contributing significantly to increase the Medicare.

This has hurt physicians as well as hospitals, physicians in large measure because physician time with patients is what is not adequately compensated. I don't want to sound too Orwellian here, but this is a date that's been bandied about in the press a few weeks ago has a possible date when Medicare bankruptcy sort of – this may be greatly exaggerated and this date is likely to bounce about over the next several years. Nonetheless, all these sort of Armageddon kind of statements and images form the backdrop of a discussion that we're having today. And, as we earlier remarked, if this were an inexpensive test, we wouldn't be meeting here. And, NIA, CMS and none of us would be here. We would say, "Go ahead and do this." But, this is the reality, heavy work.

Most of our board, most members of our organization believe that and this statement has to be sort of prefaced by saying, given the current state of the science in the field, given the current efficacy of the cholinesterase inhibitors and the NMD

{modulators} that's currently available, which have some effect in specific patients but nothing dramatic. Most people in the Alzheimer's business and in geriatric psychiatry and neurology would probably endorse that, that there may be beneficial effects in some cases, albeit limited.

What patients need is more physician time, if you will, more caregiver time and resources to navigate the ups and downs of life every day, rather than much more scan time. However, that doesn't mean there's no role for neuroimaging and there's no role for PET scan. That's not at all where I'm coming from. I'm trying to sort of provide a perspective as to where we are. In order to evaluate reimbursement for any scan, one should think about its possible impact on diagnosis, impact on therapeutics, at least agents that are currently available, and possible impact of clinical care. This is I think where it's likely to have – should be available under special circumstances is our position, largely in the area where's there's diagnostic uncertainty, very close to what was articulated earlier for the Alzheimer's Association. And, I won't get into who's to decide how we're going to decide. Nonetheless, there may be specific instances where a PET scan can actually clear up a diagnostic dilemma. And, we had several PET imagers, researchers this morning, Drs. Frey, Reiman, Foster, etc., talk about specific instances. And, they – I know, and that's likely to occur.

The worry we have is very comparable to the one that was earlier stated, that has to be restricted access in a circumscribed group of patients where there are genuine issues about diagnosis, as opposed to a confirmatory test in all cases. There should be

documentation of the diagnostic dilemma, documentation that the clinician, in certain

specific circumscribed specialties, those with training in the area, have been through this

sort of a diagnostic algorithm and is confused about diagnosis and believes that a PET

can actually help resolve it, that areas of hypometabolism, patterns of hypometabolism

can help clarify a diagnosis, and the diagnosis is likely to result in sort of perhaps better

care, whatever the nature of the care, not just pharmacological intervention. But, it may

be placement, it may be caregiver issues, it may be driving, but meaningful

documentation of this. The idea is not necessarily to drive up the bureaucracy and sort of

introduce regulatory overkill but to avoid a kind of a blank check kind of situation,

whereas it is today for many other laboratory procedures, where all a physician has to do

is rule out stroke, and it's automatically done in a medical center. Some kind of scientific

clinical oversight, again, the nature of which sort of would be the subject of another

conference.

   And, we're very concerned about possibility of repeat scans, but one can make a

case in, you know, isolated instances, perhaps, where a repeat scan in six months or a

year is necessary. We need to be very careful. We are worried that the danger of a

blanket approval is that it'll lead to sort of a proliferation, rapid proliferation of sort of

PET scans around the country, that it will lead to excessive use, perhaps more at major

medical centers, though these things are infectious sometimes, and it spreads quite

rapidly. And that given the current state of what Alzheimer's patients need, that

resources are perhaps – considering it's all a finite pool in many ways, that we should carefully think about allocation of resources.

These are recommendations, if you will, based on where the science is today. As has been pointed out by Claudia Kawas and others, if a new generation of therapeutic agents become available and, you know, that may happen in six months, it may not happen in six years, undoubtedly, the debate will change, because the realities on the ground would have changed. And, then, if a more precise diagnosis can lead to far more effective agents that are very helpful in AD cases, perhaps toxic in non-AD cases, when even more precise diagnosis is necessary for clinical care of ongoing management, the entire debate could change. But, this is where we think we are today and, consequently, these are our recommendations. Thank you. That's part of the AA recommendations that Gary Small sent to me about two weeks ago.

<u>Susan Molchan, M.D., National Institute on Aging</u>

Well, thanks to Dr. Buckholtz' stringent timekeeping, we're a half an hour ahead of time and can start our final discussion early. Why don't I start to see if the people from CMS have any specific questions that they think we haven't addressed, for starters, you know, from their original list that they want to just kind of get out in the open. Is there anything, Steve or (inaudible)? Do you want to start with those or do you want to just open – ? Okay. I have a general question, but does anybody have any? Norm, do you want to start?

Norman Foster, M.D., University of Michigan

These are questions and comments regarding the presentations this afternoon. First, I think it's clear that we don't understand the sensitivity and specificity of the clinical exam, and we heard that that was a very important determinant. We're trying to decide whether PET studies would be available to patients in the community, and we have very little information about how accurate diagnosis is in the community. And, it would be very helpful if we had some of that information.

The three studies that Ron talked about all were studies within the context of the trial and trying to test the accuracy of diagnostic criteria and, so, they don't actually meet the important feature of how community physicians diagnose and how accurately they diagnose patients.

My second comment was that inappropriate use of imaging is not unique to PET; in fact, I think that that's extremely common, for example, with MRI scans. Frequently, I see patients who have had, you know, three, five, 10 MRI scans to evaluate the same problem. And, so, I'm very conscious of trying to order tests when it's only appropriate, but that's not going to be easy to resolve. And, to remind people that saving dollars on PET does not necessarily mean that equivalent dollars will be spent in other aspects of dementia care, which I think would be very helpful if it were.

Third, a comment on Claudia's presentation: There's a very important study that she didn't mention, published by Dr. Middleman, regarding the value of social

interventions and supports for families. And just refer people to that study, which shows

that there can be a significant delay in nursing home placement with appropriate social

guidance. And, in fact, the point that's important with regard to Claudia's presentation

is that it showed that these social interventions were much more effective when they

occurred early rather than late in the course of the disease. So, I think this is probably the

strongest piece of evidence we have that early diagnosis has real value in outcomes and

in cost savings.

Female Voice

In that study, Norm, was early diagnosis facilitated by functional imaging?

Norman Foster, M.D., University of Michigan

(Inaudible/off microphone) no, it isn't.

Susan Molchan, M.D., National Institute on Aging

Or, is that a critical –

Norman Foster, M.D., University of Michigan

(Inaudible/off microphone).

Susan Molchan, M.D., National Institute on Aging

Okay. Anybody else have any questions or comments? Kirk?

Kirk Frey, M.D., University of Michigan Hospital

Let me just follow up what Norman had started. There are a couple other things that I think need to be considered that are not captured by some of the modeling that we've seen in terms of clinical decision-making. There certainly should be a penalty to the making of a false positive diagnosis of Alzheimer's disease. One aspect of this is certainly the expected side effect, downside of cholinesterase inhibitors. But, a far more important consideration is that when cognitive complaints and symptoms are attributed to Alzheimer's disease incorrectly, then the search for the actual diagnosis, often a treatable one, stops. There needs to be, incorporated in these models, a consideration of what detriment occurs to the patient from not having a diagnosis of depression appropriately reached, which is an eminently treatable alternative.

The other thing that strikes me with regard to the two models we saw this afternoon, the one from MGH and the one from Duke, I think it's worth pointing out that neither of these models embodies the recommendation now made to CMS about the selective application of PET. As I interpret the design of these studies, and I'm prepared to be corrected if wrong, these were considering the use of imaging in all comers and in a context where the imaging result trumps the clinical wisdom. That's not what's being proposed to CMS, and I think that, certainly, if you had a clinical certain diagnosis of Alzheimer's disease, no matter what the PET scan shows, it would be a mistake to rely upon that to the exclusion of the remaining data.

A final thing I have to offer is that while I understand the possible cost impact of an additional diagnostic study being reimbursed by CMS, what we're really talking about is a cost shift, in that if 40% of subjects evaluated receive an errant diagnosis of Alzheimer's disease, when indeed it is not present, in the current market, what we're doing is asking those patients to bear the expense of $1,000 to $3,000 a year worth of pharmacotherapy, which comes from their pocket. If we had an entirely nationalized health system, that expense would be charged against making the proper diagnosis, and I think if you looked at the cost analysis done by Silverman and his group, there's a considerable cost savings of around $2,000 to $3,000 per PET, if you assume the range of diagnostic accuracies that we've been talking about. So, in toto, the healthcare costs per case of Alzheimer's disease are substantially reduced by making the diagnosis more accurately.

Susan Molchan, M.D., National Institute on Aging

Scott, and then Mike.

G. Scott Gazelle, M.D., Ph.D., MPH, Institute for Technology Assessment

Yeah, I can address some of your concerns about the modeling. The first one was the penalties for that, and our model at least did include that, which is that we modeled not only the side effects of therapy, but the course of disease of all the non-Alzheimer's patients and so that if the diagnosis was incorrectly made that they had Alzheimer's disease, they would not have been treated for whatever other conditions they had, based

196

on the assumed case mix, whereas if it was not, they would have been treated. So, our model does take that into account.

Second, our strategy was that the neuroimaging test would be used only in the setting of possible, probable AD, not when the diagnosis was made definitively on the clinical exam, so only when it was considered likely but not definitive. I mean, well, it's a question of what that is –

Male Voice

(Inaudible/off microphone)

G. Scott Gazelle, M.D., Ph.D., MPH, Institute for Technology Assessment

Okay, but – well, no, but so, then, the thing is what does "uncertain AD" mean, and how do we know what the relationship of that diagnosis is, that is, an uncertain diagnosis of AD is to reality, because there is no such study. So, that's a limitation of our knowledge, and it's easy to change, though, because all you do is change the sensitivity and specificity. That's basically what we did, and we ranged that through huge ranges anyways. And, so, the modeling studies answer that question using different words, that is. What you're suggesting is simply a change in the sensitivity and specificity of the clinical exam.

And, then, the last thing is the cost to the patients. You know, when these analyses are done to guide policy, the recommendation is that they should be done from a

societal perspective and then, of course, it doesn't matter who's paying for the therapy. But, we do recognize were you to do it from a patient perspective, the results would be somewhat different, just to clear some of the things around the model.

Male Voice

The Alzheimer's Association recommends that Medicare reimburse for PET under very limited circumstances, but the concern is that if such a recommendation were taken and it was allowed – Medicare started paying for PET – that it would be great overuse of this, and it wouldn't be limited. My question to the people from CMS is: Based on the experience with other diagnostic and treatment situations, what tools do you have available to you to monitor, to limit, to deal with a situation such as this?

Steve Phurrough, M.D., MPH, The Centers for Medicare and Medicaid Services

Our tool bag is fairly limited. We provide, in most of our coverage decisions, fairly definitive guidelines on how that technology that we are providing coverage for should be used and, in most cases, assume that the reasonable providers are going to follow the criteria, recognizing that that, in fact, doesn't happen in a whole host of cases. The only real tool we have is sort of the retrospective look that various punitive agencies may do sometime in the future and then go back and recollect some money. IG can do that; there's some fraud investigations that can do that. But, in general, that's done on such a small basis that that's not very limited.

Our preference in this particular issue is, again, to attempt to, in some manner decide how we can use the authority that we have to provide the coverage under a trial basis. And, you know, we'd like to continue to work with NIA and the Alzheimer's Association and so forth to see if, in fact, we can do that. If we define it under a specific clinical trial, then you have to be part of that trial to get reimbursement, and there are coding systems that we have in place to do that. So long answer to your question is: Our ability to control providers' use of technology is limited only by their conscience.

Male Voice

I swore I wasn't going to say anything again, but I just compelled – (laughing). What made me think about this was the comment that technology may have some cost shifting. I've been around this business for about 30 years. I've not known where technology has actually reduced costs. There may be isolated examples, but what technology does is it does, at times, create demand, and as you create demand, you increase your variable costs, and as you increase your variable costs, you do end up having larger healthcare costs. Now, I'm sure there may be examples around the country where technology has reduced the cost per unit of service, but I worry about, on a diagnostic side of the equation, whether or not something as we advance our technology, unless we have that ideal drug that you were talking about, where we might have some significant impact on stopping or reversing the disease, I'm not sure that technology cost shifting is where we want to go in this discussion for the future of diagnostic imaging.

If I could make a few final comments, this I think has been very helpful to us. We appreciate all the work that Susan and Neil and NIA did in putting this together. We obviously are not clinical experts in the field of Alzheimer's or in the (inaudible) field. Our expertise {isn't in} the methodological issues. I think we're fairly good at looking at data and attempting to draw some conclusions on that, and it's very helpful to hear what some of that is today.

As you've heard, we have a very specific request in front of us now to provide coverage for Alzheimer's in a fairly narrow population versus the very broad request that we denied last year. We would like your assistance on that. In your little PowerPoint blue folder, you have this document. That's the CMS name in the top left-hand corner. I don't know where that logo came from. It isn't the CMS logo, but it's there, and it's on our website – (laughing) – so I don't know where it came from. Is it? Really. Shows you where I am.

In here, we have defined the request. We've listed the input we had from the Alzheimer's Association, and we have a series of questions, some of them clinical and some of them policy. Well, you are the clinical experts and, so, we would request that you assist us in attempting to answer these clinical questions for us. Some of them are policy questions. We would like you to sort of move from a clinician to a policymaker, and if you were the CMS banker, how would answer these questions? And, the page has

a place for you to send in your answers to, so you can email them to us. So, we would solicit your help in attempting to answer both the broad question of how we should provide this coverage, as well as some of the more narrow questions around if we did broaden our coverage, here are some of the implementation issues.

And, then, the final comment is we do want to continue with this trial design issue. We'll work with NIA, perhaps on some other forums. We, in fact, may have a forum at our office in the future where we'll invite a similar crowd to join us to see if we can continue to move forward with what the trials, that these are currently underway that we can modify or a new trial that we could put together that would let us answer the questions and be broad enough that we can introduce PET scan into the appropriate populations and learn some stuff at the same time. We think that's important, and we would solicit your help in doing that.

Again, Susan and Neil, thanks very much. This has been very helpful to us.

Susan Molchan, M.D., National Institute on Aging

I have one final question myself. You've been approving PET scanning for lots of different cancers over the past few years, but I believe that's only after it's been shown that it makes a difference in treatment outcome? Is that right?

Steve Phurrough, M.D., MPH, The Centers for Medicare and Medicaid Services

Our standard is that if it makes a difference in patient outcome. We rarely see that. The trials don't do that in general. Our second level is we like to see that it changes management of patients. We change drugs. You change protocols. We see very few of that and, so, we in fact are left very often with sensitivity/specificity data, which commonly is not done. Comparative, it's done in isolation. And, unlike Alzheimer's, in cancer, SUVs are extremely – "critical" is not the right term, but discussed significantly and there are a whole host of issues that are not standardized around PET scans in cancer. And, so, we have a lot of negative decisions in PET scanning (inaudible). We have six underway right now, and we have a number constantly.

Susan Molchan, M.D., National Institute on Aging

So the ones you've approved, you base those decisions based on a change in patient management?

Steve Phurrough, M.D., MPH, The Centers for Medicare and Medicaid Services

In most cases, our approvals are based on patient management. There are the rare ones that looked at outcomes. In most cases, it's patient management, and in a couple of cases, we had (inaudible) consider (inaudible) data, so we were comfortable that the (inaudible) trials (inaudible).

Susan Molchan, M.D., National Institute on Aging

Yeah, so just wondering if what Dr. Matchar, that Dr. Matchar's talk and Dr. Gazelle's talk where we saw, you know, no difference in treatment outcome today, if there's any precedent and if something, you know, as kind of nebulous, even, as the value of knowing would be enough, you know, of a criterion, how much that would be useful in basing an approval or how much that would go into making a decision or how valuable that would be. I mean, it's can be valuable to some people, not everyone, and –

Steve Phurrough, M.D., MPH, The Centers for Medicare and Medicaid Services

Well, you know, I felt like Claudia's position there, we have a great value in knowing as physicians. I'm not sure that there's as yet that good, clear evidence that beneficiaries have that same value of knowing. So, would that make a difference? Perhaps. We would like to see that physicians are changing management. In changing management, that we know makes a difference, and that's perhaps part of the difference in oncology is, in general, when you change management, you have a good potential for making a difference when we change management. We're not – that's not real clear yet, so (inaudible) tremendous difference in changing management in Alzheimer's.

Susan Molchan, M.D., National Institute on Aging

Right, that's right. Okay. Well, anybody else have a comment or a question? Okay, well –

Neil Buckholtz, Ph.D., National Institute on Aging

Thank you all for coming – (inaudible/off microphone). (Applause).

Steve Phurrough, M.D., MPH, The Centers for Medicare and Medicaid Services

I'm sorry, one last thing I meant to say. I will be sending this, since I have your names and email addresses, I'm going to be sending you stuff. And, if you're not on the list and would like to get stuff, you know, let me know. And, once I send one, if you don't want to get it again, let me know, and I'll take you off the list. But, I want to sort of continue this sort of group dialogue if we move forward in this.

(End of Meeting)