Howard Fillit: We'd like to get started. Okay, thank you for your attention. Good morning, I'm Howard Fillit, Executive Director of the Alzheimer's Drug Discovery Foundation and a member of the ACT-AD coalition. On behalf of ACT-AD, I'd like to welcome you to the second AD Allies Workshop on Clinical Meaningfulness in Alzheimer's disease.

Our thanks go out to the FDA for helping to make this event a reality and for all of the vital work you continue to do in this area. In particular, I would like to recognize Dr. Russell Katz, Dr. Celia Witten, Dr. Terry Toigo and Captain David Banks, and others here today from the FDA.

I also want to thank our co-host from the Alzheimer's Association and the leaders engaged in Alzheimer's disease organization, LEAD. As well as all of the members of the ACT-AD coalition for their support of this forum.

For those of you who are not familiar with ACT-AD, we are a coalition of 50 national organizations representing patients, caregivers, researchers, employers, consumers and other health-related groups committed to bringing interventional therapies to Alzheimer patients within the next decade.

Our goal is to work collaboratively with the FDA to ensure that the growing Alzheimer's epidemic is met with a sense of urgency and that the Agency has access to the cutting edge scientific information and tools it needs to assist them as they review new therapies for Alzheimer's disease.

Today's meeting will build on progress made during a similar forum held last year on the topic of clinical meaningfulness. I was pleased to be a participant as well at that earlier meeting. At that time, the FDA clarified its current standard of evaluating clinical meaningfulness in Alzheimer's disease treatments.

We heard presentations from many of the same clinicians gathered here on what factors need to be considered when determining clinical meaningful benefit in treatments for the earlier stages of Alzheimer's disease. After these presentations, there was an expression from all stakeholders that establishing how clinical meaningfulness is defined and measured across all stages of Alzheimer's disease is more important now than ever before, and we are encouraged. And pleased that the FDA has agreed to explore this topic further with us and we look forward today to a

productive exchange of ideas, and to get the ball rolling, I'm going to introduce Rob Egge from the Alzheimer's.

Robert Egge: My thanks to Howard for laying out this meeting so well for all of us. I am Robert Egge, the Vice President in Public Policy and Advocacy at the Alzheimer's Association and it's great to be with you. Thank you very much for coming today, on behalf of the Alzheimer's Association.

As you may know, the Alzheimer's Association's vision, boiled down, is basically what we all share, which is a vision of a world without Alzheimer's. And as we know from Austria and other context, we're not exactly sure what that road is, but we do know it passes through Rockville and that the FDA is critical to this conversation. So we thank you very much, in particular, for taking your time this morning to join with us in this working meeting.

From the Association's behalf, we are determined to advocate several things, and one is that the FDA is able to join their tremendous talent with 21st Century tools and technologies, all the very best to help you do your work, that's one thing we're focused on. Another area that we've been focused on is exactly the context of a meeting like this one. Whether it's here or research roundtables, wherever the context, this kind of work to clarify what we can about the development pathway. As we know, industry is making investments that are substantial and putting quite a bit at risk, just like academic researchers are, in terms of their careers and their time, and whatever we can do to clarify what that path is and reduce uncertainty, we think is gonna be very helpful to getting to this destination that we all share.

So with that brief welcome in mind, I'd like to turn it over to George, or excuse me, to John Dwyer on behalf of LEAD, for his welcome as well. Thank you.

John Dwyer: What Rob didn't know is my mom always wanted to name me George. My name is John Dwyer and I have the privilege of being one of a large number of folks that are active in an organization called Leaders Engaged on Alzheimer's Disease. We represent 30 or more organizations that work very closely with the Alzheimer's Association, ACT-AD and other constituents groups in the Alzheimer's effort to create a mega community that is trying to identify areas of need across the whole ecosystem of interests and requirements to move forward to finding an ultimate therapeutic solution to Alzheimer's.

As we look at where we are today, there have been some backdrop events that give context to this very important meeting. First, I'd like to be sure to thank Dan Perry and Howard and ACT-AD for catalyzing this meeting, because the willingness of the FDA to once again engage on this very important topic can only be facilitated by those of us outside of government, who can help them bring these kind of gatherings together, to the Alzheimer's Association for their continuing support for bringing these kind of initiatives together. And I can say on behalf of all of us at LEAD, we will do whatever we can to make sure these kind of exchanges can occur in a collaborative and transparent way to move this forward.

I did want to say two other things about the context in which we operate, because as many of you know, just recently – and Rob was a fundamental part of that, as was Meryl Comer, who is in the back of the room, the Alzheimer's Study Group just issued a national strategic plan. I encourage everyone who has not yet read it to leaf through it, because there's some very big ideas in that plan that raise important fundamental issues for how all of us in the broad constituency worried about this disease might think about moving forward.

Of particular note was their observations about the critical role the FDA plays in whatever comes out of a fruitful execution of that plan. I might add, since the FDA's been blessed with new leadership, that when Dr. Hamburg testified before the Senate, she made two recurring points. One was her deep, unbounded commitment to safety but also a recognition that she had to work with the rest of the FDA team to strike a balance between innovation and regulation.

There are a number of us in this room that when we aren't listening to people a lot smarter and frankly doing very important and productive work than we do, we're over a few miles down, around Capitol Hill listening to healthcare reform and the debate that's going on in that context. And while they're trying to reorganize over a $1.5 trillion industry with an enormous likelihood of success some day but maybe not today, there are at least 435 separate opinions about how that reorganization should occur.

I want to tell you that every day I go out there I am struck by one thing, there is consensus around one point and one point only, and that is what happens here at the FDA in conjunction with industry and science is the one consensus upside they see in changing the cost curve, in changing the quality of life for Americans for

generations to come. So this is not an academic exercise, this is not just another meeting, this is where the hard work of health reform really will occur, and a lot of eyes are upon us as we try to move forward with a new and different transparent way of reaching common ground on these kind of issues.

So with that, I want to again thank both ACT-AD and the Alzheimer's Association, the FDA, all of you for your service and for having this event today. Thank you.

John Morris: So, I'm John Morris. I'm a neurologist from Washington University in St. Louis, where I direct the Alzheimer's Disease Research Center and I'm going to be leading the first discussion panel. I want to also thank ACT-AD and the Alzheimer's Association and all involved in organizing this meeting, and particularly to Dan Perry for his persistence and Cynthia Bens for her organizational skills.

One of the organizational skills that Cynthia showed us last evening was pulling us together for a working dinner and it helped us sort of set certain touch points as we begin today's discussion. And the discussion was actually, ably led by Jeff Cummings. As always, Jeff was very wise. He waited until everyone had had at least two glasses of wine so the likelihood of consensus was increased, but I think we came up with these sort of guideposts, if you will, about today's discussion.

First, Alzheimer's disease almost certainly is a chronic illness of many years, many decades duration and the symptomatic stage, the dementia stage, likely, is the end process of that long stage. We have decided, however, that the emphasis for today will be on the symptomatic stage of Alzheimer's, not the pre-symptomatic, not the pre-clinical stage, but there is general consensus that the earlier we are able to identify the symptomatic stage, called insipient Alzheimer's, prodromal Alzheimer's, mild cognitive impairment, very mild Alzheimer's, the greater the likelihood of success of therapeutic agents. So the focus today is on the earliest symptomatic stage.

Second, at least in research settings, biomarkers, particularly imaging and fluid biomarkers, increasingly are accepted as representing the pathology of Alzheimer's disease in living patients. So we have been using biomarkers in research for some time. We think that as we talk about the earliest symptomatic stages of Alzheimer's, we have all been impressed that rates of

decline are very slow. It takes a while further down the symptomatic slope before rates of decline begin to accelerate.

In part, that slow rate of decline may be as we move earlier in the diagnostic scheme to very early symptomatic disease, we may be bringing some individuals into that sample who do not have Alzheimer's disease. They may be normal, and hence, they're not progressing at all. So the use of biomarkers in selecting samples of early symptomatic disease may help increase the likelihood that the individuals recruited into the sample actually have the underlying illness. So we advocate to the use of biomarkers to help identify samples.

As we'll hear later, it's also possible that biomarkers may indicate which individuals in the sample are more likely to deteriorate at a faster rate; that is we may be able to enrich samples with more rapid progressors.

And third – the third touch point that I wanted to mention is cognition, certainly is perhaps the essential biomarker of disease. It's the individual's declining cognitive abilities previous to his or her earlier cognitive abilities that is the essential feature of Alzheimer's disease. So intra-individual cognitive decline is a key biomarker and certainly should be measured as we evaluate the efficacy of new agents for Alzheimer's disease.

But for individuals with the disease, declining cognition is important only as to how it predicts the ability of that person to function in their everyday life. I've been evaluating and managing people with Alzheimer's disease for some 25 years and I've never had a family member come to me and say I am so worried my husband had dropped two points on his mini mental state score. Families don't really care about the level of cognitive status unless it affects the ability of their loved one to function, to do everyday activities, to prepare a balanced meal, to shop appropriately, to handle household finances, to drive an automobile, to operate a television remote. These are the key issues for clinical meaningfulness, and so we also advocate that we have not only a measure of cognition but a global measure that captures everyday function.

And with that, we are going to hear from our first presenter, Ken Rockwood from Dalhousie University in Halifax, Nova Scotia, who is going to be telling us more about how to measure individual function and decline. Ken.

Ken Rockwood: Thanks, John, and thanks, too, to ACT-AD for the opportunity to speak today. So, as you've heard, my charge today is to talk about global, but also about individualized outcome measures. So I have some disclosures at the outset. Some of the pharma companies, but importantly, to this company that I've set up to help assist in how individualized outcome measurement can be made more feasible. It's sort of a Canadian thing, standardized individualization. I have a bunch of references for what I'm gonna say. I'll leave them here. I'm happy to talk about them at any point.

So, I want to start by painting the picture of early Alzheimer's at the symptomatic stage and MCI in a very particular light, and that is to say I think there are two important challenges that we face. One is, this is a multidimensional state. So in the definition of dementia, that's clear, global, progressive, cognitive impairment that interferes in social or occupational functioning.

But even in MCI, I think that as we look carefully at what people complain of, we understand that there's more to even AMCI than the amnesia, people have more symptoms than that, and that is a challenge for how we think about and measure in this area. I think that's fairly noncontroversial.

A second important challenge that we face, which comes out very much in the MCI end of things is dynamics, and I'm speaking here in a particular technical sense, about the way that things change over time, and the dynamics of cognitive change in older adults has a very specific signal, which we've seen in more than a dozen datasets with a variety of cognitive measures. So it's not dataset dependent, it is not measurement dependent, but let me just show you one.

So if we look over in this panel, where it says N=3, that's the error group of people here from the Canadian Study of Health and Aging, these are data on 9,000 people looked at over five years across all cognitive states. And N=3 means that they made six or seven errors on the 100-point 3MS. And you can see that across the range of the 3MS there is a very characteristic change in the distribution. The data points here on the Y axis, this is simply the portion of people who have a particular score, and on the X axis, it's the scores that they have. The line is the model, the points are the fit to that model. And so you can see this is a highly fit model and this is change over five years.

So someone who started off in the three error state, the chances are that after five years, about 55 percent of them will have progressed to have more errors. Very few people have a very large number. The mode still is that they'll stay at about three, but that's only about 18 percent of people. And there is a chance that there will be improvement.

And this idea that at any given error state, even though on average there's loss, some people get better is very important, because this tells us, or so I claim, that cognition and cognitive change, even with Alzheimer's, is not the brain as an innocent bystander. The brain fights back and there are ways in which it fights back and times in which it fights back, in ways that we don't entirely understand. So we have a picture of things apparently having either stabilized or gotten somewhat better for reasons that we don't understand.

And so this has been a challenge to the area, because when we look at this clinically and say what does that improvement look like, in MCI terms, it's sometimes been thought as simple diagnostic instability, and even when it's accepted something real has happened, people use these sort of slightly pejorative terms of well, they've regressed or they reverted or there's diagnostic backwash here. And my claim to you is that this is an important challenge for us, because I see this as being intrinsic to what the state of things is, and there will be a bit of a tradeoff as we think about biomarkers and picking out the people who are the most likely to progress, there will be a tradeoff in terms of what we understand the generalizability to be.

In the dementia field we see this too, and we've now got three large population-based studies that have been done over two to five years, showing that about one person in seven to one person in eight, apparently is about the same stage over that time. For some time, we thought that this might be good vascular risk factor control but that doesn't seem to account for it.

So we need to understand that the relentless progression model of cognitive change will be an incomplete account. We have to have a way that embraces the complexity that some people are going to get better. And I expect even with very highly powerful biomarkers, there will still be some improvement and we should think that that's an important thing to expect.

So my claim is that global measures are actually a very good way to take into account both the dimensionality and the dynamics

which are important challenges. There are a number of standardized ones. I think a very widely used one, about which we know, is the CDR sum of boxes. Obviously, they tackle the dimensionality, they don't just measure cognition. They measure the impact in a variety of other areas.

And the other thing is, they help to measure the dynamics as well, at least they help to smooth things out, because the frame of reference is not just what happened today at test time, it's been put in a context of weeks to months of change.

But importantly too there are individualized measures, and this is something that I've spent some time trying to work through, and my claim to you is that they do about the same thing. And I would put in this the CIBIC-Plus in its original intent, not sort of the more ad hoc measures that it sometimes has evolved into, is a way to take into account what's important to the person in front of us.

John just now said that what was important was not cognition but other things, and then he listed off several of them, driving a car and cooking and all sorts of stuff. Well, some people don't ever drive a car, so that will never be important to them, some people don't cook, that will never be important to them to get that back, they never lost it. So, individualization is a key way to get at the signal here is my claim, but it has to be adjudicated by an experienced observer is what our findings have been.

So now I want to come to the heart of it in terms of what it means to individualize, and the Oxford English dictionary makes it clear that to individualize is to make particular to that person. But the consequence of this is that not everyone will have the same considerations in terms of what's important to them.

So when you compare people in terms of an individualized outcome measure, they will each have their own basket of goods. Some people will come and their issue will be social conduct and other people will come and their issue will not be. What you need then is to standardize the extent to which we say they've met those goals, no matter what those goals are. But this is an important conceptual point, because I view it as the single biggest strength that we have there. We're gonna find out what's important to you. Some people view this as a terrible weakness that what's important will not be the same for each person. As it turns out though, most people agree on some general patterns, so it's not quite chaos.

The rationale to use individualized measures then is multifold and it speaks to the validity. Obviously, if you've improved someone's particular goals, then on its face, the clinical meaningfulness question is easier to get your arms around. It's also important from the generalizability standpoint. You don't need to adjust for culture or education. People will choose goals or choose areas of concern which are important to them. And so it helps from the generalizabilty point, the clinical meaningfulness point.

The responsiveness issue is a technical, mathematical, statistical thing and that's empirical. I think what we've found so far is that the individualized approach is as responsive or more responsive than the standard ones that we have had. It helps from the knowledge translation standpoint. You can go from the trial to talk to family doctors, to families and say here are the things that appear to have gotten better here. Here is the sorts of things that you might want to get at.

And importantly too, it has this advantage that allows us to see in a trial what we didn't know to look for at the outset of the trial. And particularly, as we move into new therapeutic agents, which may work in ways that we hadn't anticipated, the ability to see what we didn't know was there, I think, will be very important and will allow us to escape from simply being anecdote based.

Let me give you one example. This is a randomized, placebo-controlled trial that I did in Canada, it was called the Vista Trial, The Video Imaging Synthesis of Treating Alzheimer's, because we videotaped every single clinical interview that was done there. At the outset, we did not know to look for repetitive questioning. Even though it turns out in that study and others that we have done, it's typically the most troubling of the common symptoms. It's the ones the families give the most priority to. I can't dad asking me this same question 30 times a day. We didn't know to look for that at the outset.

In a secondary analysis that we published in 2007, here is what we found, that of the 50 people, with 120 in the trial, the 50 who gave a priority to this, 14 or 58 percent of the people on the treatment group met their goal for improvement in repetitive questions. Compared with 20 percent of the people who took placebo, only 8 percent of the people at the end of the double blind phase at four months were worse compared with about 32 percent of the people who were on placebo. So this was an important signal that we didn't know to look for despite having worked with the Cholinesterase inhibitors for some time.

We used a technique called goal-attainment scaling as the primary outcome. We had one arm of it done by the clinicians who were involved, the treating doctors. And then one other bit was done by patients and caregivers and the CIBIC-Plus rater rated or helped set the patient-caregiver goals or at least paid attention to that interview but didn't adjudicate them at all. At the end, both the treating doctor and the CIBIC-Plus rater, blind to each other, agreed on clinically important improvement with an effect size, that was about the same. We calculated it as a standardized response, meaning because the particular one we used here, goal-attainment scaling score and you can see that the effect sizes of the two approaches were about the same with that.

Not everybody is in favor of this. I've hinted that some people see it as a big objection, you're not measuring the same thing on each person. You're only measuring the same idea about the extent to which they have attained their goals. There's worries about whether this is acceptable to gaining. What about if you set your goals too low. That was too subjective and is it feasible and so on. So the feasibility issues are met by training and the other issues are met by the blinded-placebo control or at least controlled design because if there is goals that have been set too easily they'll be set too easily in both groups or so the claim would be.

So, I will sum up by saying this, that global measures, both standardized and individualized, allow a systematic response to these important challenges of the high dimensionality of the problem that we face and to the dynamics. I think that, in particular, individualization also offers us a very quick way to enhance our understanding of the clinical meaningfulness and as well to look for things that we didn't know to look for at the outset, which one hopes, as we get new compounds working in different ways, will be something of ever increasing value to us. And I'll stop there. Thank you. *(Applause.)*

John Morris: Ken, thank you very much. I'm going to invite a distinguished group of panelists to come up, Eric Siemers from Eli Lilly Company, Dave Knopman from Mayo Clinic Rochester, Jeff Cummings from UCLA, and we're going to discuss the role of the global in measuring clinical meaningfulness.

Just to begin I want to concur with the points that Ken made about individualization, and hopefully, Ken would agree with me that some of the global scales that he mentioned, the global deterioration scale or GDS and the clinical dementia rating or

CDR, in fact, are designed to assess individual decline, which as I mentioned in my remarks, is really the sine qua non of dementia, intra-individual decline in activities custom for that person, would you agree?

Ken Rockwood: Yes, I do.

John Morris: Okay. All right, but Ken, maybe we'll start with you. And by the way, this is open discussion, so please come to the microphone and identify yourself and make a comment or ask a question. So, Ken, you very persuasively made the case for global scale in measuring the efficacy of drugs and the global scale should focus on individual decline. You mentioned there were some challenges, you didn't dwell on those challenges. Do you want to – you had a listing of about five or six and is it too much time and so forth. Do you want to expand upon that?

Ken Rockwood: Sure. So one of the challenges at the outset is that you have trained people to use it. It's a new thing. And that's a very pragmatic issue because it means that when you set the trial protocol up you need to allow for that to be done. Because we've spent so much time emphasizing the need that a given measure be reliable, then there's a bit of a mind shift that people have to make to think about that everyone will not have the same things that are important to them. And in some people's minds, it's hard for that to ring true.

I think we found it an objection that is best overcome if you put it in the context of the CIBIC-Plus, that there's a need to individualize and to anchor. And as well, with the CDR sum of boxes or the NPI, they're the exact manifestation of a given symptom will not be the same in each person. So that offers a way in for that, but in having done this in a series of trials in Canada now, I do know that that is something that you really need to carefully pay attention to.

After that, there is also a concern about what if people set goals that are too low, and there are various ways to deal with that. As I say, from a pragmatic standpoint, the design takes that into account, that one group won't be favored over another group. And again, it's a matter of the training, that people don't set goals that are too low.

To help adjudicate that, though, you really do have to know what dementia is and how it works, so you can help people understand is this a reasonable goal for them.

Having done this for some time, I've been very impressed by the extent to which people bring a realistic expectation of this to the table. I don't think – well, it certainly has not been my experience that people come to a trial wanting to be cured – I mean they want to be cured but they don't expect it. And an interesting thing that we've observed is, just like if you were to say to someone, how much money would you need to make to be happy, most people say about twice what I make now. Very often, if you ask people to think about what would be an important treatment success for them, it's if the intensity or the frequency of the symptom were to go down by half.

And I just want to make this point in a bit more elaborate way. The standardized approach can tie us in this knot; two people come and their issue is repetitive questioning. We learned about this now. So we said, okay, when we go to have our mini mental, ask the same question test, and so we say, look, on average we've done these big studies, we know that most people can take ten times a day and it's no problem. So someone comes in with 30 times a day and they want it to go down to 15, and someone else comes in with 10 times a day and they want to go down to 5.

And so we say to the people, you know, five, well, you've met your goal, because on average five should be fine. And so we need to have a way for people to express what's important to them, taking into account the way things vary for them. I think the CDR sum of boxes achieves that too. It provides some context. If we were to tackle this, though, the worry that I would have is say in MCI, where people have identified two things being important to them, interest in hobbies and the memory bit. And say you improve both of those, now you have to divide two by the entire denominator. If people only set two goals, actually, we increased it, so it's set between four and six, so two isn't the best example. But if they did meet two things, we'd only – if they did set only two goals and they both got better we would, in that way, divided two by two. So the responsiveness, in theory, should be enhanced.

So I would think that we really do have a need to have a way to get to grips with expressing the individual desire for treatment and seeing the extent to which we meet that desire.

John Morris: So, I have some experience with the CDR and I think a problem, a potential problem is exactly the title of Ken's last slide, synthesis. So the interviews and the individual decline do require an experienced clinician to put it together and that can be expensive

and time consuming and sometimes subjective, so that's the issue that I think is another factor. Jeff, you had a comment.

Jeff Cummings: Yeah.

John Morris: You want to use the microphone, Jeff.

Jeff Cummings: The microphone, yes. Thank you, Ken, for a clearer presentation of goal attainment scaling. My own view is that goal attainment scaling is applicable when one is dealing with symptomatic agents; that is a patient has a problem and wants that problem to get better. However, I think the framework of this discussion with the FDA is that we are moving towards a class of compounds which is likely to be disease modifying and where there may be no symptomatic improvement. That a win would be slower progression or stabilization, and I do not see how goal-attainment scaling can help us in that regard because the patient does not know what function is going to be lost next that they value most.

So, I think we'll have to return to a discussion of the classical dual-outcome criteria of cognitive impact and global impact. The global has traditionally included everything, so behavior, function and cognition and our particular problem with moving this to the early stage patients is that they have very little functional impairment, that is how we define them and very little behavioral impairment or disturbance because it doesn't emerge until later in the course of the illness.

So we have only cognition. Now, my question for you, John, is can we catch cognition in the global? Is the global still an applicable concept in these very mild AD patients? How do we establish clinical meaningfulness beyond cognition? Is cognition enough? These are the things that are troubling me.

John Morris: Well, I certainly, Jeff, don't want you to be troubled, so –. I think it is difficult to capture change in the very earliest symptomatic stage of the illness, change in cognition, change in function. And I agree that generally there is not much behavioral deficit. I do think that the rate of change can be captured if the sample is most likely to have underlying Alzheimer's disease. I think some of the heterogeneity of capturing individuals who don't have Alzheimer's disease does interfere with seeing a rate of change. So it is a small change but it is capturable and I think both cognitive and global measures are able to do that. Eric – oh, Dave?

Dave Knopman: I just would echo what Jeff said and ask the question, the concern – the question is a concern that when you define globals on an individual basis and there's this variability from person to person on how stringent or relaxed the goals are, it would seem to me that that's going to increase noise and that's going to make it more difficult to show a benefit. Now, it may be washed out by the fact that you are doing it in a double-blind fashion, but if it increases the variability, that really is not suitable. So that's an empiric question, can you comment on that?

Ken Rockwood: Yes, I can. Our experience has been, and certainly our published experience has been that the variability has not been increased by this. One of the slides I cut out was a nice portrayal of how the standardized response means go. So, as you know, the standardized response mean is the mean difference between the groups divided by the standard deviation of the change and because goal-attainment scaling for that particular example is a change score that's the best way to go.

So, what it showed from the VISTA Trial, which was the placebo controlled trial, but also with two others, one called ACADIE and one called VASPECT, is that the SRMs of goal-attainment scaling have been the highest of all of the measures that have been used. So though there is a theoretical concern there, that's not how it has played out.

And I think it's for the very reason that you say, because there is a bit of a tradeoff there. To get back to Jeff's point, I think one of the ways – so I would say two things. You very concisely said that this approach might not be suitable for disease modifying agents because how would we handle unexpected new effects, and the way that it's done is those become identified as new symptoms and they are given the status that they were not set at the baseline interview of being worse than expected outcomes. So that's one way that that gets handled.

The other point, though, is in our experience with AMCI patients, it's not just memory that they complain about and often what drives the person to come to the clinic, even though they go out with an AMCI diagnosis, are things like they can't perform their job as well or they can't perform their hobbies as well, or the family finds that they have less interest in going out. They'll still go out if they are taken but they, themselves, have less interest to do so.

So I do think it allows us to individualize at very early symptoms what sometimes go below the clinical threshold but yet at the same time, actually motivate the behavior to come to clinic.

John Morris: Eric, I was gonna ask you a question and I see you have a comment, which is terrific. The question is, of course, you've spent a lot of time in considering ways to measure clinically meaningful effective agents, so I hope that you'll be able to comment of your assessment of the role of a global scale in conjunction with a cognitive measure and particularly, Ken's goal-attainment scaling. So can you mention both of those in your comments.

Eric Siemers: Well, sure, thanks, and that's basically what I was going to comment on. I thought maybe just to start off, my probably most important disclosure is that my immediate boss is Richard Mohs, and so, as many of you know, Richard was involved quite a bit in the development of the ADAS, now that's not the global measure, of course, but I think that the principle is the same.

And John, I think you can back me up on this, having talked to Richard a fair amount about this, there's always this issue about as we move to early patients, whether or not – and again, this is for the cognitive measure, but whether the ADAS performs well. And when you think about it, it was developed at a time when the field was really focused on people with more moderate disease and every scale has ceiling and floor effects. And the fact that it works at all on an MCI-like population or whatever you want to call it, I think is pretty remarkable.

But if you talk to Richard, and maybe it's better that he's not in the room here, he will tell you, he doesn't care, it's fine, if people want to develop a scale that's more appropriate for people with very mild disease, that's not a problem for him. He's done it, he doesn't get paid when people use the ADAS and he has no problem with any of that.

So, I think to get to the point about the global measure, then, it's the same sort of issue, is how do we determine the appropriate global measure in these very early patients. So you made the point early on that we were just going to talk about symptomatic patients at this point, which we need to limit it to, but just to think it through, if you think about people who are pre-symptomatic, if you have people with autosomal dominant mutations, in that case, we don't have any clinical measures. By definition, they're pre-symptomatic. And so we have to come up with some sort of

artificial thing to measure, whether it's a PIB scan or whatever it is we come up with.

As you then move into people with very early symptomatic disease, particularly when you are talking about treatments that we hope will slow down disease progression rather than just providing an immediate symptomatic benefit, in my view, no matter what you use, a lot of these constructs are gonna be a little bit artificial. In other words, whether or not you can use the remote control for the TV may be really important to me, my wife probably wouldn't think so much.

So I think we can really make the argument that when you're dealing with these mild – very mild subjects, that whatever the instrument we come up with, to some extent, it's going to be a little bit artificial in terms of how much of an effect this has on their daily lives. But what we do want to do is we want to use these tools, and they are all just tools after all, to assess, in many cases, whether the treatment has actually slowed down the progression of the disease.

And just one last point along those lines, there was a comment made initially about the cost of Alzheimer's disease, which of course are huge. Not everything is dollars and sense, but if you look at the cost, of course, they go up exponentially as you get into later stages of the disease. So we're trying to use these tools and measures at the very early stages to tell us that we're gonna have a big impact, we hope, in the later stages of the disease, even from just a cost standpoint.

So I guess the overall point is, I don't know that we need to struggle too much with the idea that this is absolutely, clinically relevant, and we can debate forever whether it's important to use a remote control, but we really have to accept the fact that these are tools that we use to assess a broader process.

John Morris: Right. So, Dave and Jeff and Ken and Eric, I'm going to switch for just a moment and ask all of you to consider the criteria developed by a working group led by Bruno Dubois that does not incorporate measures of function or behavior, focuses simply on cognition, primarily memory but incorporates at least one biomarker in the definition of Alzheimer's disease. So no global, no function, what do you think of that?

Jeff Cummings: Well, John, that's just a matter of diagnosis that doesn't really get at the issue that we're talking here in terms of measuring longitudinal

change. As I'm gonna talk about, I think that there are some substantial merits to that approach that I didn't quite realize when I first saw it – and I'll just save my comments for later. But I don't think that the Dubois criteria preclude using global measures as outcomes in longitudinal studies.

John Morris: But how about in defining the sample that was enrolled in the study?

Dave Knopman: Well, I think that – right, I see your point. I guess I was going to comment on that as well, that although it is usually the case, and several people made the comment that when researchers look at MCI patients, you almost always see that, in fact, they do have subtle deficits in function, whether it's in providing consent, whether it's in doing financial matters, whether it's in other more complex activities. And so simply by defining somebody as having an amnestic disorder, you certainly increase the likelihood that you're going to be – that in fact, some of those people on subjective or objective measures are going to have difficulties with function.

But I think still you run the risk that many of them will say, no, I'm fine, and the spouse will say, he's fine, he does everything, asks questions repeatedly but I swear he has no other problems. And I think if you start to require functional impairment in your diagnosis, you've already dropped your severity down to a level where we don't want to go now. We want to try to get it early as possible.

John Morris: Yeah, I will disagree. I think that one can define, as you mentioned, the subtle functional declines and capture individuals who are having Alzheimer's disease and capture them actually at a cognitive level prior to where they would meet criteria for MCI. So even pre-MCI, you can diagnose people accurately based upon functional loss.

Dave Knopman: I think across the board, especially in people of lower educational and occupational attainment, that that's not the case. I think if you're dealing with people with 20 years of education, yeah, I agree. But for people at lower levels, I think people who are less sophisticated, caregivers, I think it's much more difficult.

John Morris: I think it takes the role of an experienced clinician, so maybe we can –.

Dave Knopman: We can make things up for them.

John Morris: No, I think – so there's disagreement in the field, but I think you can capture people, and not with 20 years of education, who are declining.

Dave Knopman: But actually, I want to make the point, I don't think while we may disagree, that this is a fundamental problem for the FDA and for how the disorder is going to be defined. I think that – I don't think that we need to agree or disagree whether there is or isn't functional impairment in these people. We can do – follow some procedure to identify people with the earliest symptoms, however we wish to do it, and as long as we identify those people who we think are at higher risk to decline on biological grounds, that's the key.

John Morris: Right. Eric.

Eric Siemers: Just one point along those lines, I guess, sort of regardless of if you can come up with that functional measure, but I just want to comment on the point that was raised earlier about variability, especially in the rate of decline. I just think it's interesting to think about that, people who possibly revert back to, if we want to use the term MCI, MCI to normal or don't progress, if you put those people through the filter of having some sort of biomarker diagnosis. In the ADNI dataset, at least, last time I looked, of the people with MCI who had had spinal fluid analyses, there were five people, I think it was five, who went from MCI and reverted back to normal. Well, all those people had normal spinal fluid.

And I think one of the things that the biomarker aspect of diagnosis will help us with is to reduce the amount of noise. And so I think we can definitely cut back on that.

John Morris: Right, right. Jeff.

Jeff Cummings: So you asked specifically about the Dubois criteria. I think we're gonna come to that in the last session also, but because you asked, John, I'll just make the point that the Dubois criteria do not limit the diagnosis to people who have only memory impairment and no functional impairment. They simple do not require functional impairment and the are equally applicable to late-stage disease as to early-stage disease.

The unique application of the Dubois criteria is it allows you to make a diagnosis of Alzheimer's disease before there is Alzheimer's dementia. And I think that's where we're trying to

move. And they do that by the use of biomarkers, which is a strong position.

John, I wanted to push you a little bit on CDR, that's clinical dementia rating.

John Morris: Thank you.

Jeff Cummings: You've been a great champion of that instrument as am I, and I think it might well be the leading instrument for capturing global changes in a mild population. But I am, as a champion of behavior in Alzheimer's disease, I've been bothered slightly by the fact that the CDR does not include behavioral measures, and therefore, by using the CDR as your global, it excludes one whole dimension that the CIBIC, for example, would include.

John Morris: Mm hmm.

Jeff Cummings: So, I wonder whether you could respond to that and is there some way that we could count on behavior and still use the CDR?

John Morris: So, excellent point and much like Eric mentioned about Richard Mose, and the ADAS or the ADAS Cog, the CDR certainly is simply a tool. If it is useful to capture cognitive and functional change, then it can be used, but it is certainly not the only or necessarily the best tool for all purposes. So, one of the deficits is that it does not measure behavior very well. Another deficit is it does not specifically measure language. So, these are issues, the CDR is incomplete.

The advantages perhaps over the CIBIC, is at least it does have a standard training system and reliability and certification process so that it can be compared across trained users of that instrument. But I would come back simply to say that if there is a better instrument for the purpose at hand and the CIBIC serves better, then that's certainly fine with us.

Ken Rockwood: You mentioned the NAC instrument that has incorporated the other two –?

John Morris: So, the uniform dataset is an instrument that is used by all 29 NIA funded Alzheimer's disease centers to provide a standard clinical and cognitive assessment of individuals who are healthy controls and individuals with MCI and mild Alzheimer's disease, so that all the data is obtained in a uniform manner and the data are maintained at the National Alzheimer's Coordinating Center at the

University of Washington as a great repository to stimulate collaborative research. So this uniform dataset includes a number of the measures that we've discussed, including the CDR. So the CDR is the clinical, global measure of this uniform dataset.

Because of its inability to capture behavior, the uniform dataset also includes the neuro psychiatric inventory, the NPI, and because the clinician judgment for language and behavior in particularly for disorders other than Alzheimer's disease that might be encountered in the dementia setting, there is an extended rating of the CDR specifically for behavior and for language that are being used to capture that. And Dave was helpful in providing those extensions.

But the CDR is the global instrument for the Alzheimer's Disease Center and the uniform dataset. Eric.

Eric Siemers: One quick comment about that, just since I'm the pharma person up here, is that most of us have trials that are multinational at this point, the large trials especially. And it's my understanding, and maybe some of my colleagues can comment on this if they've heard differently, but at least from an EMEA standpoint, the functional measure that they will require will be something like the ADCS/ADL.

Now, personally I sort of like the CDR. I think it captures maybe not everything but it captures quite a bit. But I think from their standpoint, they felt that maybe it was a little bit too broad, but in any case, the guidance that we've seen is that the functional measure, they would not be in favor of using the CDR as a functional measure.

So one of the things that I think is up for discussion is since the goal of this is new treatments, that if the phase three studies really are multinational, we have to come up with instruments that are going to be acceptable by multiple regulatory agencies.

John Morris: Right. Yeah, I cannot comment about that. I know that the CDR has been translated by drug companies into some 60 languages. So it must be – that doesn't preclude using the ADCS/ADL score, I mean, but that's right.

So I think we have a few minutes, if there are any comments from any of the attendees, will you come – Rachel, come and use the microphone please, introduce yourself.

Rachel Schindler: Rachel Schindler, Pfizer, New York. I agree very strongly with a lot of the issues that Ken brought up about the need to individualize things, particularly in early AD, where – particularly if you're talking about somebody – you know, David mentioned, at that higher educational, more sophisticated level, where it becomes really difficult, may become difficult with our instruments to detect decline. And certainly, we all know, we're not really probably gonna see functional decline, except as really reported by an individual who can't do maybe a very, very high level activity. Particularly in say, a younger person in their early 60s who is still working at a very high-level job.

The biggest concern that I have is in the issue of less than expected decline, which Jeff kind of touched upon in this issue of disease modifying or symptomatic, but I don't think that that even matters, because I think that even in the symptomatic – the so-called – and I don't like the distinction but we want to use it for a minute and I'll use it for a second and then put it away –. Even if you use the symptomatic treatment paradigm, meaning that the effect of the treatment works – that the treatment works while you're using it and it stops working when you take it away, we're still seeing, certainly in our datasets, and we all know – all of us who have treated patients, less than expected decline, which the goal-attainment scale, which I love, but still – I'm not sure how it accounts for that.

So, we're faced with the problem that, so in large datasets where we're comparing placebo to treated, we can detect the difference even if both groups are declining. But it becomes virtually impossible in a single individual patient sitting in front of you, which is the – maybe not a regulatory issue, but I think ultimately it all gets wrapped up together. Because that's probably the biggest reason that people don't get treated, because they see deterioration, and again, whether that's with – that could happen with a so-called disease modifying agent or with a so-called symptomatic agent, patients gets worse but less than expected decline is huge, especially based on the conversation last night. I think it was Ron was talking about – or maybe it was Eric, talking about the five-year – you know, the longer term is really what we need to look at but we really can't for practical reasons in clinical trials.

If in five years there is deterioration but it's so much less than it would have been because you're maintaining that much more function, I think everyone would agree that that's the important. So I think just coming full circle, I'd be very interested to hear, and particularly Ken, your comments about the less than expected

decline issue. And particularly as you're talking about a drug that could be used in everyday practice.

Ken Rockwood: Right. So, less than expected decline is a statistical issue, right. So, what would happen if neither group declined very much on the measures that you use, then the effect size would be small and it would be hard to detect an effect when you use a big sample size in order to overcome that.

I think that what the panel has proposed is that biomarkers would be used to have an enriched sample in which less than expected decline would be an unexpected thing. And after that the drug has to work, the drug has to actually show that people who take it don't get worse over the timeframe studied compared with people who are on placebo. So that's a challenge for whichever test is put forward.

So the empirical account then becomes is the sensitivity to change of this measure or of an individualized approach better or worse than the sensitivity change of the other ones that we use. In the symptomatic trials that we've up till now, with a maximum duration of one year, it has been the most responsive or as responsive as any of the standard tests.

So if the claim is that we've got a drug which will work but in a very subtle way that people can't tell, you know, that's gonna be a tough measurement issue no matter which measure you use.

Rachel Schindler: See, I would – I guess I would argue that we're dealing with a neuro degenerative disease, which by definition is progressive. And when I said less than expected decline, I don't mean small amounts in both groups, what I mean is that the person declines less than they would have without the treatment.

Ken Rockwood: Oh, yeah, yeah, sure. So the way that that would be captured is people are asked to think not just of how things might get better but how things might get worse as well. And so they would say, let's take the repetitive questioning one, they might say, right now it's ten times a day, if it went to five times a day that would be better and if it went to zero that would be great. And they might say if it went to – you know, if someone is at the very edge, they might say if it goes to 12 times a day, that's it, I can't take it anymore. I'm at the limit of what I can stand. So for them, that's what the goal is. For someone else, it could be 15 or 20 times.

So people are asked to imagine what is a frequency that would be somewhat less than an expected outcome or a much worse than expected outcome. And so after that you simply aggregate over the people in the trial to understand where the range of the treatment affect is in a quantitated way. So people are asked to imagine degrees of decline.

Rachel Schindler: So, again, you're assuming that the individual can project a trajectory.

Ken Rockwood: Yes.

Rachel Schindler: With some amount of accuracy. We probably need to get on to something else, but I'm not – I feel sort of like it's still not – to some extent, you're still – somebody doesn't know what that trajectory is. But I'm not meaning to argue. Just trying to solve a problem.

Ken Rockwood: Sure. Okay, and so I won't argue back except to say that most people who come to a trial have experience with the patient they care for and how that person has changed over time. And so they're able, within the bounds that we've seen, to forecast how they imagine it might play out.

John Morris: Okay, so I was going to ask Dr. Katz to comment on the role of the global, and particularly the idea of an individualized goal-attainment scale, but he is at the microphone, so –.

Russell Katz: Yeah, hi, Russ Katz, the FDA. Yeah, we have certainly endorsed – I don't know that we've actively encouraged, but we've certainly endorsed the use of individualized measurement in lots of different settings, not just in Alzheimer's, but certainly also in Alzheimer's. No one yet really seems to have taken up on it. I gather there are multiple issues and there are things that people are concerned about.

But from a regulatory point of view, the fact that different people have different functions that they value and you can't necessarily compare across patients or across studies, is not – is in the way, a bar to the acceptability of relying on individualized globals. I'm heartened to hear, I think, at least in certain clinical settings, a consensus perhaps that a global is still important even in the early patients. And we have endorsed that principle, however it's measured, and so far continue to endorse that, absent any compelling evidence that we shouldn't.

I did have a question for Dr. Rockwood. Maybe you could talk briefly about who gets to choose what function is important to whom. We're always concerned that a drug may have an effect on a patient that is beneficial for the caregiver but not necessarily for the patient, and you could imagine repetitive questioning is something that bothers caregivers, and one could imagine a drug which is sedating or, you know, produces a depressive affect that diminishes something like repetitive questioning which might be appreciated or valued by a caregiver but not necessarily by the patient. So I wonder if you could just talk about that sort of thing, particularly related to degree of disease, so in an advanced patient, maybe it's the caregiver who has most of the input, but in an early patient, where the patient themselves still have some insight, maybe they choose. So if you could just talk about issue then.

Ken Rockwood: Sure. Well, I think you've summed up exactly what our experience is, that as the disease gets more advanced, that it really is the caregiver who drives this. Earlier on in the process or over particular goals on that patient's value, they commonly have input. As I said in the Vista trial, we videotaped every single clinical interview, and it's actually very poignant to look at how those interactions go, what input the patients have.

You can imagine there are some patients who don't get a word in edgewise and one estimates looking at a series of these over the eight months of the trial, that probably is a lifelong account of how things have gone. But I must say, it was very heartening thing to see the care paid to the view of the patient, even when there had to be a certain degree of prompting to understand exactly what the patient was saying. It was very impressive to me to see the extent to which the caregivers quite commonly are able to interpret for the CIBIC-Plus – or sorry, for the nurse who is doing the facilitation of the interview.

The patient might say two or three words and the caregiver could make an inference about what that meant and was very keen to say that. So there is a certain variability. Our experience has been that people tend to pay attention to what the patient thinks, even as the disease would progress more.

The repetitive question one is an interesting one, though, because virtually never does a patient complain of that. They may assent to it being possible when it has been identified, but I don't think I've ever seen someone in which the patient said this is an issue for me, I've asked this same question again and again.

John Morris: All right, well, listen, will you join me in thanking our distinguished panelists. *(Applause.)* And particularly Ken for his presentation. And now we turn it over to Dave Knopman.

Dave Knopman: Well, thank you John. I'd like to also thank ACT-AD and FDA for allowing this meeting to take place and us to have this discussion. I'm not going to do any speaking now but just turn it over to Jeff, who is going to talk about the issue of clinical meaningfulness and – no, that's interesting but wrong one. It is Cummings.

Jeff Cummings: Thank you, David. Thank you, ACT-AD, thank you, FDA for being here. My charge is minimally clinically important differences, how little is too little. I'm going to make a very specific recommendation. I'll invite Dr. Katz and Dr. Black and Dr. Knopman to tear it limb from limb, and maybe that will be a constructive exercise.

My disclosures are that I've consulted for many pharmaceutical and biotechnology companies, including those represented here today. I'm gonna touch on symptomatic therapy and what I think our lessons learned from there were. I'll look at some disease-modifying therapy experience, try to apply the lessons from cholinesterase inhibitors. To those, I'll develop a 25 percent rule. I'll discuss the special circumstances and caveats of the application of that rule.

Why begin with cholinesterase inhibitors? Well, they are approved by the FDA, they're widely used by doctors, they are funded by payors, they're widely taken by patients. Most patients are on them. They're a general agreement that the effects are small and the provide a reasonable place to begin the discussion of the lower bounds of what we would consider as effective therapy.

I have chosen for my discussion the ADAS Cog, because it is the most commonly used outcome measure in clinical trials, and I have chosen to begin with mild to moderate AD, even though our emphasis in this meeting is MCI, because it is the most common trial population and treatment group with which we have experience. I'll then try to apply the learnings from this group to mild patients.

Here are two reviews cholinesterase inhibitors, looking at the effects on the ADAS Cog. There were ten studies in the Cochrane Review by Birks. He found a range of effects of 1.4 to 3.9 points on the ADAS Cog. In a 14 study meta analysis that was published in BMJ, the range was 1.5 to 3.9, almost identically, almost

identical. I derived, therefore, that 1.5 ADAS Cog points is a lower bound for the effects of cholinesterase inhibitors.

This is what this would look like. I don't have a pointer. It's usually a six-month study. You see that that's from 2 to 2.5 there on the bottom scale. Most of the studies are six months long. We infer from data, we do have one two-year double-blind placebo controlled trial that we have something like parallel decline after the initiation of therapy. So that's what that therapeutic response looks like.

Now, can we take these lessons to disease modifying studies? Well, most of the studies so far have used 18 months and have thought of the ADAS Cog decline as being 6 to 7 points in 18 months or 4 to 6 points for one year. There is nothing special about 18 months but it is a common exposure in a disease modifying trial.

So the recent experience with Tarenflurbil shows 7 points in 18 months and with Bapineuzumab 8 points in 18 months. Here, for example, is the Tarenflurbil study. A failed drug but a successful study, because the placebo group declined in exactly the way we thought that it should. You see that at 18 months there, we have a 7 point decline and at 12 months we have a 4 point decline. So I think it helps us to understand what to expect of the ADAS Cog in an 18-month exposure study.

So this is what 18 months looks like and that's 6 points, and then if we added our cholinesterase inhibitor line, it would look something like this. So if we looked at the 18-month point and said that a disease modifying drug should meet the lower bound of success of a cholinesterase inhibitor, we would reach this point, a 1.5 point ADAS Cog drug-placebo difference at 18 months, which is a 25 percent difference from the expected 6 point decline on the ADAS Cog. That's what you would get if you choose this lower bound.

So that's a 25 percent drug-placebo difference. If we use the usual standard deviation of three times the difference, you would come up with an effect size of 3.3, that is considered small by the usual classification of effect sizes.

So here is my proposal for a minimal clinically important different, that it be 1.5 drug-placebo difference in mild to moderate AD, that is statistically significant. Alternatively, one consider that same conclusion to be a 25 percent drug-placebo difference at 18 months, that is statistically significant or an effect size of 3. Those

three things are all the same for an 18-month mild to moderate study.

So the 25 percent rule and 25 percent and statistically significant rule could be applied in other situations, for example, with more mild populations. You could still work with an effect size of .3 in a pre-Alzheimer's dementia. The pre-dementia phase of Alzheimer's disease. You could use it with other tools, for example, with a neuro psychological test battery.

I note that this suggestion is somewhat different than the 2 point ADAS Cog, 18-month drug-placebo difference recommended by the European Consortium of Clinical Trial Investigators. That was reached by consensus with undeclared assumptions and opaque process, but it – nevertheless, it's a point to be considered.

On the left you have the NTB, in mild disease and in moderate disease, and you see that it tracks nicely in both phases of the disease. On the right, you have the ADAS Cog, showing very little change in mild disease and good change in moderate disease. You can see that it would be easier in this particular set of projections to show a 25 percent change with the NTB in early disease than with the ADAS Cog in early disease.

Here is the ADCS/MCI trial. I'm showing you just the ADAS Cog component of it. I have outlined in red, 18 months and 36 months. You see that at 18 months in this MCI trial – and this was a true MCI trial, which means that it was not specifically chosen by biomarkers to have patients who had AD only, and therefore, I think our point will be, as we get to the next session, that we can do better than this.

But even choosing this very heterogeneous population, you see an ADAS Cog decline in the placebo group of 1.29 points, a 25 percent difference would produce a drug-placebo difference of .32 and again, the effect size would be .35. At 36 months, you see, now we have 3.74 points decline in the ADAS Cog, 25 percent difference would be .93 and the effect size would be .31. So these general rules look like they could be applied in an MCI population as well as in a group of patients with mild to moderate disease.

So, I think that the 25 percent and significant rule or an effect size of .3, said to be .3, not 3, is a reasonable expectation. The rule is empirically based on clinical trial experience and makes very conservative assumptions. A 25 percent drug-placebo difference would allow new agents to compete with generic Donepezil after

18 months of therapy. For the first 18 months, you'd be just as good, you would function just as well on Denepozil as you would on a disease-modifying therapy if the disease-modifying therapy had only this effect.

We assume, but it would be up to the industry sponsor to prove that longer-term therapy would be associated with increasing drug-placebo difference.

Why did I choose such conservative assumptions? Well, I think our goal is to facilitate drug approval. We want new drugs for patients with Alzheimer's disease. We want to encourage industry to engage in AD drug development and this is treacherous territory, expensive, and so far, there are only failures.

We want to encourage the development of agents with novel mechanisms of actions, where the effect size may be unpredictable and are likely to be small to begin with. And finally, I think that this very conservative approach to begin with would initiate the process of incremental improvement in disease modifying agents, which is how I understand the process of drug development to proceed. I'll stop there.

Dave Knopman: Now, if we could ask Dr. Katz and Dr. Ron Black to come up. I think I might say, just while we're getting started that we will probably not go fully till 10:15, maybe till 10:00 and take a break. So to start with, I wonder if I could ask Dr. Katz to comment on Jeff's presentation and how the notion of an effect size fits in with FDA law and the approval process.

Russell Katz: Well, as you probably know, we don't have any a-priority requirement for a particular effect size. We, as the purpose of this conference is, we have relied on the concept, however one defines it, of clinical meaningfulness as being the thing that we really care about when we approve a drug. Does the drug have an effect that's clinically meaningful.

We didn't really know when we set out these rules, and I'm not sure we still know today, what a minimally important effect size on a particular scale, and certainly not on a measure of cognitive function. That's why the global was decided as being necessary as a co-primary. We said that any effect that you see, however small, on a specific cognitive measure, if it can be detected to be clinically meaningful in some reliable, appropriate way, that's good enough.

So, there is no requirement really for a specific effect size on any particular scale. We either choose a scale – and I'm speaking now, sort of generically across indications, we either pick a scale that we think that any change on is clinically meaningful or if we're not sure what a clinically meaningful change is on a particular symptom scale, let's say a cognitive measure in Alzheimer's disease, then we impose a requirement that there be another scale that documents the effect that you've seen is clinically meaningful. So the short answer is there is no requirement in the law for a minimum effect size.

Dave Knopman: Okay.

Russell Katz: I guess I have a couple of questions for Jeff.

Dave Knopman: Go ahead.

Russell Katz: It's not clear to me that – what necessarily the purpose is for defining a minimally effective treatment effect on something, let's say like the ADAS Cog. Is it designed to obviate the need for an independent measure of functioning or global functioning? Is that the purpose, so that we can say here is a minimum effect size that we think will perforce translate into a clinically meaningful effect, and therefore, we don't have to measure a global? Is that the point?

Jeff Cummings: No. I think that could be part of the understanding. I am thinking more the reason to do this is because in designing a trial, we'll need to have certain power assumptions and that will be related to the anticipated effect size of the drug. And in advancing a drug, one would need to decide, well, how little is too little, because with a very large sample population, you might be able to achieve an effect on the ADAS Cog that's statistically significant but we would, I think at least as an academic community, agree that it wasn't clinically meaningful.

So I was trying to define something that was clinically meaningful but I'm not sure I would go so far as to say that by itself it's clinically meaningful.

Dave Knopman: Ron, do you want to make any comments or ask Jeff questions from industry perspective?

Ron Black: Sure. Well, I can speak from the perspective or working really almost exclusively in disease modifying therapeutics for a number of years. And here, you know, the target that – I mean the way you

do this is you – first, you target pathology. So you try to find a way to reverse the pathology of Alzheimer's disease and hopefully, if you're theories about the pathology are correct then you end up with an effect that's clinically measurable.

Now, the problem is, is that, I mean I think, Jeff, as you pointed out very well, is that early in the disease, you're not actually measuring very much change, and in fact, the pathology itself, which is the target of our therapy, begins 10 or more, as John Morris would say, years before the disease begins. So it's difficult to put an actual target in terms of the change that we're looking for. Because it may be that there's a lot of pathologic progression in an early stage disease that results in this much difference from placebo, whereas a relatively small change in pathology later might result in that kind of a difference when the scales are more sensitive, for example, in the more moderate disease.

And I just – I worry a bit about some of these barriers, because I mean I think eventually, to get to truly disease modifying drugs, we're gonna have to take a sort of more encologic approach, and look for treating the right patients and finding complementary therapies, because this is really proving to be a hard target in terms of modifying the disease.

Dave Knopman: I'd like to ask all of you, since it is the case that it's not a quantitative measurement that FDA would use for the approval process, maybe this issue is more relevant in the bigger picture in terms of payers and in terms of what society is willing to pay. I wonder if you guys could comment on that. Do you think that there is an effect size that a large insurer or pharmacy benefit company, maybe - Dr. Katz, you can't make that comment, you're certainly welcome to, but Ron and Jeff, where they would balk? Do you think that this issue – how relevant do you think this issue is at that level, not at the approval process of FDA but rather, at whether they're gonna reimburse for it? Jeff and Ron?

Jeff Cummings: Yes, well, that's one of the reasons that I was thinking about this concept is, I think if a new drug is not at least as good as our existing drugs, and the existing drugs are very cheap, then I think it's gonna be very hard to promote and have payers buy the new drug. So that's why I looked at what the effect of Denepozil would be and then what it would take to have a drug be at least as good as Denepozil for at least the first 18 months of therapy.

And I think as a field, we're still blind to how much persistence we can expect from our patients. We know that with the current

symptomatic therapies, persistence is certainly less than a year. I hope that that is not the optimum persistence. If it is, then of course, disease modifying therapies will have little benefit, because the whole point of disease modification is that they are taken for a long period of time with the expectation that the drug-placebo difference will get larger over time. So if patients won't take it then we won't be able to expect that benefit. So, I'm thinking that having a bottom – a floor, is important from a payer perspective.

Ron Black: Yeah, I mean one of the problems is, you know, when we make arguments to payers trying to support the pricing of our products, I mean the most compelling arguments are directly related to cost. And in fact, the cost in Alzheimer's disease accrue mainly later in the disease. For example, in an early Alzheimer's disease population, there's very few of these high-cost endpoints like nursing home placement or loss of ability to take care of one's self.

Whereas, again, in the early stages of Alzheimer's disease, again, where we would most like to target the pathology, I mean we're talking about things like being less able to use the remote control. I mean certainly some of the patients are working and might no longer be able to be in gainful employment, but most of the differences in function that reflect the pathologic changes are not really changes that have – very directly translate to costs. So it makes that argument with payers rather more difficult.

Unless you take a longer view – you know, longer meaning even beyond the duration of your trials towards what the eventual cost savings or quality of life savings will be to your therapy.

Russell Katz: Well, I, of course, can't speak to the question of what's an adequate enough effect size to motivate a payer to pay, but I do want to say something about effect sizes. As I said before, we have no a-priori effect size in mind. We try to pick scales so that any statistically significant difference on those scales, whether it's one scale or several scales, can be interpreted as conferring a benefit to a patient that matters to them. So there's no number that we look at.

But I have to say that implicitly, in some sense, we do think about that, because the reality is that most trials that we see are sized about the same as other trials we've seen, which means that they're powered to pickup a treatment effect of the sort that we've seen in the past. If somebody came to us – now, of course, that's changing with disease modifying agents, because I'm not sure we know what to expect and there isn't much experience with those.

But if someone came to us, let's say, with the next cholinesterase inhibitor and proposed a 30,000 patient study, we'd balk at it. And although that's a study that's power could pick up infinitesimally small treatment effects on both – I mean if that's what this treatment does, this hypothetical treatment, so you could detect a statistically significant difference on an ADAS Cog and any global you choose and it would meet our current tests. But I think we'd say, wait a minute –.

So there's an implicit view that treatment effect is sort of back there, you know, in the back of our minds. But what it is quantitatively – if someone came down with a study – if someone came with the next Alzheimer's drug and it had a slightly smaller effect on the ADAS Cog and the global than the previous treatments but it was statistically significant in a reasonably sized study, we'd probably say okay. But as I said, if it was 30,000 – if you needed 10,000 patients to show an effect on both of those outcomes, we would definitely balk at it.

Dave Knopman: So there is some sort of smell test-?

Russell Katz: Well, you could call it that, I suppose. But there is, I think, an implicit – we bring to the table an implicit understanding that it's got to be reasonable, whatever that is. So, you know, I applaud Jeff's attempt to define what's reasonable, whether that's the right standard or not, I don't know, but we think about that or it's there in the back of our minds.

Dave Knopman: Ron, did you –?

Ron Black: Yeah, I mean I don't want to defend 30,000-patient trials, but we're running some big ones. But one thing about – in a different paradigm, that is the paradigm of a disease progression therapy, there's an extra added wrinkle. Not only are the patients on cholinesterase inhibitors, that's one wrinkle, but if your drug is actually not going to improve cognition, if it's really, hopefully only going to stabilize cognition, then your trial is really dependent on the deterioration of the placebo patients. And I think there that's a harder thing to predict. I mean certainly we've seen widely varying placebo deteriorations within clinical trial sites, countries, clinical trials in general.

So, I think that just the uncertainty that comes from having to wait for the placebo patients to get worse, when your therapy is not actually making people think better, I think results in some pretty large trial sizes.

Russell Katz: Well, sure, as I said, in the next phase of development, where we're talking about disease modifying, we don't really know what to expect, necessarily, we don't know how many – I picked 30,000 – I don't know if you're doing a 30,000-patient –. I didn't mean that as the number above which you can't get approved.

My only point is that we do think about treatment effects or, you know, there is sort of an implicit understanding that a study that's grossly overpowered is a problem.

Dave Knopman: Right. Your boss once, ex-boss once said, well, you could always buy statistical significance.

Russell Katz: Yes, *p* values are purchasable. They are if your drug is effective and you can exclude bias, fraud and chance, as they used to say. But that's true, if your drug is trivially effective, you can power a study sufficiently so that you pick up an effect and we, of course, don't want that.

Dave Knopman: So if anybody has any questions, please – Dale.

Dale Schenk: Yeah. I think there's something that nobody's quite said this.

Dave Knopman: Just, introduce yourself, just for the –.

Dale Schenk: Oh, sorry, I'm Dale Schenk, I'm with Elan. Anyway, it is something that nobody has quite said, I think, is that clinical effectiveness is highly pent on the customer, whether the customer is the patient or the customer is the caregiver, whether the customer is the regulatory body like FDA, or the customer is society. I think they're all, each individual.

So for example, and this is all hypothetical, but if a drug has a very small effect on ADAS Cog, but for one reason or another radically affects co-morbidities and dependency of the caregiver and other problems associated with say diabetes or whatever, hip fractures, that's extraordinarily important to society, of minor importance though, to say the sponsors and other customers. I just wanted to get a reaction to that comment. So there's the approval issue and then there is the what happens in society issue, and they're two very different things I think.

Dave Knopman: Mm hmm. You guys want to –?

Jeff Cummings: Well, the only thing I would say, Dale, is that I think we're trying to use clinical measures to somehow get at different angles on the pathology of AD, so that the cognition is seen as, you know, a major outcome of the pathology and behavior and function, and they're all windows. And it's conceivable that a drug would have more effect on one of these windows or be more apparent in one of these windows than another.

But we'll still have to meet certain standards to get those drugs approved, even if that's the case. So I think we come back to the standard measures.

Russell Katz: Yeah, I think in order to get a claim as an Alzheimer's treatment, we expect, of course, that the drug will have an effect on those symptoms that are specifically referable to the Alzheimer's disease. And it's, as you know, our current standard is some cognitive effect and some functional or global functioning effect. If a drug happens to have effects on other co-morbidities, if it's an antidepressant, sort of independent of this Alzheimer's effect or has some other effect on some of the other co-morbidities that's worth exploring, I suppose. But again, and this is old news, but to grant an Alzheimer's claim it's got to be Alzheimer's specific, whatever that means.

Ron Black: Yeah, but just to highlight again, I think this gets – the earlier you get in the progression of Alzheimer's disease the harder this gets, and we've been working on trying to develop sensitive scales in early Alzheimer's disease for a long time. And whereas variability is a problem in standard sort of mild to moderate Alzheimer's disease, if you think about some of the subtle deficits that occur in early patients, as well as the variability of the cognitive and functional demands on these patients, it gets quite difficult and it's hard to peg an actual, you know. Again, it's laudable to go after goal-attainment scaling and things like that that are individualized but it gets very difficult because the spectrum of deficits is just so broad in those early patients.

Russell Katz: I would think that the individualized, global measure might be more useful in a case like that, where there is great variability from patient to patient about what they're functional deficit is, if it's tailored to the individual patient, that just seems from a logical point, maybe it's not true, but just from a logical point of view it seems like that would be attractive.

John Dwyer: John Dwyer again. In my day job I have the benefit of sitting on some early stage venture backed companies' boards, as well as

dealing with some of the – both the domestic and foreign regulators with respect to reimbursement. So I have a question but make two quick observations. First, I want to compliment Jeff on his effort, because I think one of the challenges you have, whether you're big pharma or a venture-back startup is, you're so hungry at the Board level for some kind of baseline decision-making framework to have some understanding about whether the scientists, the regulator or the reimbursement environment, are we in the zone of likely success or not so likely success.

So the model is – I was in the back saying, boy, that's the first time I ever saw anything that good in front of me about, even directionally, the likelihood of a product coming through a clinical study in an acceptable way. So my question would be on that front, to you, Dr. Katz, is that kind of a direction something that you can work with at the FDA to give us, as a constituency, just a pathway to explore with more clarity and certainty? Because I think Jeff's approach, that's what – I don't know if I agree with the baseline drug your testing against, because overseas that may or may not be such a great baseline. But as a directional approach, is that something you folks would be comfortable with exploring more vigorously?

Russell Katz: You mean in terms of looking at what's a minimally important difference and moving forward from there? I mean as I understand the point of what Jeff has presented, it's really to sort of power studies, to decide how many patients you should enroll in a trial. So we're certainly always interested in that question. As I said, we don't particularly have a number but we are interested in not detecting extremely small and therefore clinically trivial effects.

So, we certainly entertain this sort of approach. But I think, again, it's to my mind, the real critical question is how do you measure clinical meaningfulness. And it still sounds to me like there's a general view here that measuring global function in addition to measuring cognitive function, no matter how early the patients are, at least considering the patients, I think we're mostly talking about today, early Alzheimer's or MCI, whatever you call it. I think that's the overarching principle that I'm hearing still seems to be, being endorsed. So sample size is something we can always talk about, but I'm not so sure sample size has been the real problem up till now.

John Dwyer: Yeah, the last time I took a biology class Darwin just came back from the island, so accept that as –. But it seems one of the benefits of Jeff's approach to a layman is that you're getting

something other than a placebo baseline to look at as you measure change in your trial that gives those of us that are wrestling with this problem from the other vantage points, a frame of reference, because of what Ron Black said about the vagaries of the placebo group. And maybe that wasn't what you saw as one of the benefits of your approach, Jeff, but it certainly was to me, and that's just worth more exploration from where I sit.

Jeff Cummings: Great, thanks. No, I was trying to provide a framework for deciding how little is too little, and that was one of the things that we're continuing to struggle with even though I'm hearing very clearly that a statistically significant difference is probably going to be enough, even if it was less than what I proposed as being too little.

But I think from the wider perspective, there are still constituencies that would see less than the effect of a cholinesterase inhibitor as being too little.

Dave Knopman: Howard.

Howard Fillit: I'm wondering if we could get a little granularity around this. So suppose there was a drug that, let's say, was an anti-amyloid therapy, that at the end of the phase three showed substantial effects on CSF biomarker, maybe a PIB scan, maybe hippocampal volumes but the global was quite marginal in a mild to moderate population. So, the thinking in the field might be, well, you know, to get an effect we have to treat people who are changing at a relatively rapid rate, so we have to go into the moderate population perhaps, but they're too late.

So we always end up in this conundrum, you know, that maybe the reason mild to moderate fails is because we're treating too late, but if we go too early we can't see clinically significant changes because the rate of change is so slow. So the sum game here might be the more global interpretation, it might be to resolve that conundrum to look at the whole picture, the whole profile, the dossier of the drug at a committee meeting or wherever that decision might be.

So, to expedite things in this world where we have tremendous unmet need and not a lot knowledge about validated targets yet in humans, you know, Rusty, how do you think one of your committee's might interpret that kind of study, where again, just to summarize, you had, let's say a big biomarker effect on a validated target like ameloid, but a quite a marginal, maybe even in Jeff's

framework, you know, marginally statistically significant effect, that in the absence of biomarker data, the committee might say it's not enough on the global change but interpreting that in the context that we have, it might be, you know, sort of provocative, let's say.

Russell Katz: Yeah, well, it is provocative and – we've – all right, a couple things. We've told sponsors that from the point of view of gaining a disease modifying claim, we'd be willing to look at a package that included a trial that had clinical outcomes of the usual sort, presumably statistically significant by the usual rules, and biomarker changes.

Now, if the clinical changes in such a study were marginal, let's say –. And again, it's hard to talk hypothetically, but you know, if the ADAS Cog was statistically significant and the global just missed and you had very impressive biomarker changes, you know, that's certainly something we would consider. And we've always said, by the way, that that's the kind of thing we would take public to a committee.

So I don't know what the committee would do, I have a guess, but you know, that might be compelling. From the point of view of ultimately being able to rely upon these biomarkers in earlier and earlier disease where the clinical symptoms aren't yet quite as manifest, we've proposed anyway, as have others, a procedure whereby we start with the condition where we think we can see a clinical effect. But again, as you point out, the later, maybe, the less likely they are to be useful, and show some sort of a relationship between the change in the biomarker and the change in the clinical and then move further back where you think there's not much clinical yet to measure.

And so move sort of backwards from trying to sort of quasi validate the biomarkers where we can see clinical change and move further back so we have some reason to believe that the biomarkers change in the correct direction. But if a study that wasn't exactly statistically significant on the global, let's say, had very robust biomarker changes, that's certainly a package we would look at seriously. What the outcome would be I don't know, but we would take it public and we would seriously consider it.

Dave Knopman: Either of you, Jeff, Ron?

Ron Black: Well, again, I just think it's important to remember that we haven't yet taken the first step towards an effective treatment for disease modification of the pathology and the drug that you're talking

about might be an appropriate first step. Obviously there would be a long way to go, and again, this might be a situation where there might be synergies and combinations, for example. I would hate to not take the first step because of some of these technical issues.

Dave Knopman: John.

John Morris: I just wanted to return – John Morris from Washington University in St. Louis, just wanted to return to a small point and see if what I remember is also consistent with what everyone else remembered, and it comes to the issue that I think Ron brought up about the vagaries of placebo decline. And there are many caveats, the earlier we go, the less the rate of decline. I think we all agree with that.

But didn't Lonn Schneider give a survey of all clinical trials that had been both in the treatment era and pretreatment era and found that rate of placebo decline depended very much on the sample size. The larger the sample size of the individuals who were placebo, the more likely the placebo was to decline, and where the trial showed very little placebo decline, they almost always were small numbers, is that correct?

Dave Knopman: I don't remember that exactly. That certainly seems plausible, though.

Ron Black: In that paper, he was saying that there wasn't a secular trend towards less placebo decline over the –.

John Morris: He said that but I think he also said that the studies that didn't show a rate of decline – so there was – it didn't make a difference that Denepozil was available – that didn't seem to influence it, but also that – at least what I remember, that the studies that showed very little, if any placebo decline were almost always very small.

Ron Black: Were the smallest. Well, I mean I think that this – you get caught up here in the variability. And if you're dependent on placebo decline, you don't want to do a small study because these studies are very difficult, very costly, very time consuming. You don't want to have a failed study. And we don't say in depression trials you might do six to get two or something like that. These studies are too complicated and too costly to take that approach.

Dave Knopman: I could make a comment – go ahead John.

Jeff Cummings: Well, just thinking, John, that there were two other studies that I think are worth noting. One is the one by Michael Gold, again showing that the longer the trials the more likely you are to see predictable placebo decline, so that's reasonable. The other was the Roy Jones paper showing that within the Denepozil dataset, if you look across all the Denepozil trials, there is a secular trend with diminishing placebo decline in recent years compared to the early trials.

Dave Knopman: I'd just like to make the comment that there is a confound there that trials that are gonna be smaller may also tend to be the ones that aren't as well designed, and the centers or cast dispersions that might do such trials may select patients who have been in trials before, for example, and who are professional Alzheimer's patients. This will get to the issue of selecting the right people. But I think that the quality of the trial really does make a difference, whether we're talking mild to moderate AD but especially in the earlier patients as well.

There are any other comments? This might be a point where we could take a 15-minute break. I'd like to thank the panel and thank the audience. *(Applause.)*

Jeff Cummings: Thank you. We've had two interesting sessions. The next session also focuses on patients with early disease. The field has migrated to the concept of MCI, mild cognitive impairment, which consist of a heterogeneous group of individuals, some of whom have Alzheimer's disease and some of whom do not.

As we increasingly move towards the development of therapeutic agents that have effect specifically on the Alzheimer's disease process, it becomes increasing important that the patients to whom we give the drugs have Alzheimer's disease. So our challenge is to define which patients with so-called mild cognitive impairment have Alzheimer's disease and how that will inform our trials.

David Knopman, whom you've already been hearing from, will take us through the preliminary discussion and then we'll have panel members join us for more discussion.

Dave Knopman: Thanks, Jeff. My disclosures are given here. The problem as we've been talking about is that cognitive impairment at the earliest stages of symptomatic Alzheimer's disease poses some diagnostic problems. When we see people clinically, we can't distinguish those who are destined to worsen from those who have more stable conditions. We can't distinguish on purely clinical

grounds, people who are destined to develop Alzheimer's disease versus destined to develop vascular dementia or Lewy Body disease.

The changes, regardless of whether we get it right or not in terms of the diagnostic acumen, the changes go slowly, more slowly at this stage of the disease has been shown repeatedly than later on. And I would just like to make the point that in terms of clinical criteria, it's always possible to push the criteria down to require more severe illness for entry but then you're really defeating the purpose of trying to get to the disease as early as possible.

I just want to show some data here. First, on the top there, from the ADCS/MCI trial that Jeff commented on, the CDR sum of boxes declined .4 points. This was over 12 – this was the 12-month data, and ADAS Cog was .61, and you see the standard deviations. Those are not a lot to work with in terms of detecting change.

One of the studies that made me – really woke me up to the problem just recently – hopefully was aware of it before, was this paper that was just published by Rochelle Doody and folks from Pfizer a few months ago in Neurology, where the CDR sum of boxes in the placebo group, and I think in the treated group too, declined 0.1 points and in the placebo group, the ADAS Cog actually improved and this is a problem, this is our problem.

I just make the point that the Doody study had only 41 percent ApoE 4 positives – the ADCS/MCI is 53 percent, and ADNI, the numbers I show at the bottom also has a 53 percent APOE rate. So the question that seemed to be that has to be answered is how can we enrich the sample of people who we recruit for those individuals, 1) who have Alzheimer's disease but 2) who are going to decline at a faster rate.

And the point that I'd like to make is that it turns out that those two goals are interrelated, fortunately. People who are destined to develop Alzheimer's disease are going to decline at a faster rate than those who have a nonprogressive condition.

Another piece of data that I think is worthy of insight and I'm not proposing using ApoE as a biomarker here, but I think that the data is nonetheless interesting. From the ADCS/MCI trial, it was observed that those who were the E4 carriers had a much faster rate of progression to Alzheimer's disease than the E4 non-carriers, and I have some data that I can't show because it's just been

submitted but that also was true in terms of the cognitive measures, not just on this endpoint, but on the ADAS Cog, on the CDR. The ApoE 4 carriers had a much faster rate of decline than the E4 non-carriers.

So, there's something there that if we can enhance – that shows that if we can enhance who we enroll, we will see larger changes. This will get at some of the questions that have come up repeatedly earlier in our discussion of seeing change in the placebo group and having a reasonable sample size to see drug benefits.

So these are some of the biomarkers that could be used. And the point is, and what is emerging as the message here, the underlying message is that all of these in one way or another are getting at the – detecting people who have some sort of cognitive impairment but who also are experiencing neuro degeneration. That is to distinguish then, people who have some – evidently, have cognitive impairment but who are not experiencing neuro degeneration.

I'll just mention some data from PIB PET studies. There are three of them. One just came out in Annals of Neurology by Wolk. One is now e-pub in Neurology, Okello in the Forsberg one. And basically they all show the same thing that PIB positive MCI subjects are highly likely to decline and develop Alzheimer's disease within one to two years. PIB negative MCI subjects are rather unlikely to decline. So this is one issue that I think could be followed up on. PIB, very expensive, you have to have a cyclotron, hopefully some of the other amyloid imaging compounds might change this.

In terms of structural imaging, this is data from the ADCS/MCI trial from Charlie DeCarli, showing that those who had hippocampal atrophy were much more likely to decline. There's lots of data that shows this same thing but I just point this out in the clinical trial context that if you were to select patients at entry who, just on a visual rating, that's what this was, not a quantitative, computer-based rating but just a visual rating, the Shelton Scale, that those individuals with a greater degree of hippocampal atrophy experienced a greater likelihood of developing Alzheimer's disease.

Now, this slide, actually, I must ask that nobody photograph this. This is not published data, so – it's in press in neurology from our group at Mayo, but it's ADNI data. What I'm showing is an algorithm called the stand, a standardized index of brain atrophy

that's measured from a template of a bunch of normal subjects and a bunch of Alzheimer's patients, 160 of each, actually. The maximal areas of change are computed and then on an individual basis, that template, that model of Alzheimer-ness that comes from structural MR imaging, with the whole brain, not just hippocampus, is computed and you come out with a single stand score, a single number that reflects the Alzheimer-ness of the degree of cerebral atrophy.

For those who are familiar, this is different from volumetric based morphometry, VBM, which can only look at groups, this looks at individuals. And a couple points I want to make. This is really the key slide I want to show. What I'm showing is in normals, MCIs and Alzheimer patients, based on where they stood in terms of just percentile ranking on the stand, on the measure of brain volume loss versus the CDR sum of boxes.

So what you can see in the cognitively normals, it doesn't make much of a difference, at least at this sample size and the rate of change on the CDR sum of boxes, as you expect, is virtually zero. In the amnestic MCIs, what you can now begin to see is a separation and those who were at the 75th percentile, that is most Alzheimer-ish in terms of their MR, had the greatest change over two years on CDR sum of boxes. And even if you were to say just look at the 50th percentile, that's better than looking at the bottom group in terms of the amount of change that you'd see.

And the third point I want to make is comparing the MCI subjects to the Alzheimer subjects, you can see that there is an acceleration, regardless of – or virtually regardless of which level of structural atrophy one had, that there is a change from going from MCI to AD in terms of the degree of cognitive impairment. There is an acceleration that occurs. And what we want to try to do is capture those people who really – who meet the criteria for the earliest changes of functional or objective cognitive impairment, MCI if you will or whatever you want to call it, but who have the greatest likelihood of declining.

To switch to biomarker, the CSF is probably well known to many people, the well-known study by Hansson had shown that those who had an AD-like profile had a very high probability of declining – MCI subjects had a very high probability of declining to AD. In terms of quantitative data, John Morris' group, Snider, just published, showed that those who had a low or mid cortile, so that you'd actually capture two-thirds of the group, declined on the CDR sum of boxes 1.1 points, I believe that's over a year, versus

those who had the normal Abeta profile, actually got better amongst MCI patients.

Another study, basically the same findings, from Visser, from Europe, where the slope differences there are obvious. If you have an AD-like profile, in terms of your Abeta or whichever marker you want to use in this CSF, you will observe greater decline in that group on average.

And I'd like to make the point that you don't – we don't have to get it right for ever patient we enter, we just have to get it right on average on the group that we enroll, that they are enriched for people who are likely to decline.

This is CSF data from ADNI and it basically shows the same thing, that those who are in the highest quartiles of Abeta or tau have the greater rate of decline amongst the MCI patients.

So we need to talk about the problems and I'm sure there'll be – I hope there'll be questions about these issues. And obviously it's gonna have some affect on recruiting. I think the issue of what to tell people who screen negative and what to do with them is an issue. One thing you can tell them, maybe, that in fact, their prognosis is pretty good. The data certainly would indicate that.

And issue that I think is very important to the pharmaceutical industry is if these entry criteria are so restrictive what would that do to the labeling and would that limit the number of patients who would be eligible to receive these drugs and we were talking about this last night. If you only included people who had – MCI subjects who were PIB positive, if that was the requirement to receive the drug, then it wouldn't get us very far.

So my final slide is I think there's been a real advance in the last few years in looking at biomarkers and presenting and developing the data, and the good news is that the biomarkers not only identify people who are more likely to have the pathology that we're interested in, but those people who have the pathology, who are experiencing neuro degeneration decline at a faster rate and those are the people that are most appropriate to put in clinical trials. And I really do see it that we can kill two birds with one stone with this approach.

I'm not taking a position on which biomarker should be used. I certainly think that only one is necessary, because all you want to do is enrich the sample. And I – not meaning to push MR imaging,

because this work had come from my colleagues at Mayo, but I think there are a number of different approaches with MR that might make MR a very attractive, noninvasive measure. I think that CSF is equally attractive though. So I'll stop at this point, and thank you very much.

Jeff Cummings: Thank you, David. And if I could ask our other panelists to come up, John and Dale. Maybe while they're coming up I'll just make one point that there has been one attempt to codify exactly what David just said, which is the Dubois Criteria. These were published in the Lancet of Neurology in 2007. The lead author is Bruno Dubois, and he suggested that a diagnosis of Alzheimer's disease, to be distinguished from Alzheimer's dementia, can be made in a patient who has a memory impairment and a positive biomarker consistent with a diagnosis of Alzheimer's disease. So that's one possible approach to defining Alzheimer's disease, using the kind of biomarkers that David has suggested.

So, I'll just ask each of you to comment. David, maybe I'll start with you, I think you have something to say about the Dubois Criteria, and then John and Dale, if I could ask each of you to comment on the role of biomarkers in clinical trials, design and entry.

Dave Knopman: Well, I think that what Bruno and Feldman and Scheltens did made a lot of sense. I think, though, that the key issue is that it goes beyond simply defining the disease at an earlier state. But as I said, the use of the biomarkers identify people who are more likely to decline, and those are the people that we want to identify.

Jeff Cummings: Mm hmm, very good. John.

John Morris: So, last evening I promised to behave and not to be provocative, so I'm going to not be provocative. But I am going to say that I think, as Dave just mentioned, the issue of decline for that individual, and as Ken presented this morning, earlier this morning, is critical. I was a coauthor – original coauthor in the Dubois paper and I asked to be removed from the masthead because they used a static measure of cognitive loss, memory loss as the criterion and it doesn't capture whether those individuals who fall below whatever arbitrarily defined cutoff there is for impairment, whether for that individual it represents a decline. Nor does it indicate – and of course, they get around this with the biomarkers, with which I agree, nor does it indicate if people truly are impaired, what the cause of that impairment may be. It just says memory impairment defined as one standard deviation or 1.5 or whatever.

And I think as Dave pointed out in his presentation, that's why samples of MCI tend to be heterogeneous. There are individuals in there who perform poorly but are not declining or there are individuals in there who are impaired but not because they have Alzheimer's disease. I mean there are lots of reasons people can have mild cognitive impairment.

But if we want to improve the likelihood that an individual has been impaired because they have intra-individual decline, decline from their previous level, and the etiology is likely Alzheimer's disease, one can do that without use of biomarkers. So one can do that and be correct 92 percent of the time. You don't have to have the biomarkers, but you must – well, not must, but if you only use a single cut point on a static test, a single test, you will capture people who, 1) are not impaired, are not going to decline at all, and 2) who have – are impaired but are impaired because they don't have Alzheimer's disease, for other reasons. And that's why all of the MCI data usually show, what, 30 to 40 percent of people don't progress or don't have PIB positivity or something of that sort.

You can get through that just by clinical means alone, just by this intra-individual decline principle. I realize that that isn't the way the large majority of the samples are defined, so I do strongly advocate the use of biomarkers and I couldn't be more supportive of that, I just wanted to put that out there that you don't have to use biomarkers to be accurate.

Jeff Cummings: Dale.

Dale Schenk: Well, yeah, I was unable to make it for the dinner last night. I wasn't able to get in in time, so I made no such promises like John did to not be provocative. So I can be provocative. So I'm gonna – hopefully, I'd like to point out one or two things.

So, in general, in terms of the use of biomarkers, I just wanted to clarify one point, we all know it but I wanted to clarify it for the audience, and that is, you can use biomarkers to figure out who has or has not – doesn't have the disease, as I think, John, you were just describing. They're not absolutely necessary for that. But a bit of what you were just talking about, Dave, was the fact that you can use them, however, to select who might respond to clinical trials and specific therapeutic agents.

So, a couple of points. One thing that nobody's quite said is, well, we sort of have defined the field in terms of testing agents on –

based on symptomatic agents, and we're now moving in to disease modifying and progression modifiers, hopefully. And those are based on pathology, they're not based on neurotransmitters. And the pathology, by definition, is either Abeta or tau, we could talk about cerebral vascular disease, we can talk about synaptic dysfunction, whatever you want to talk about, but the different therapeutic agents are targeted to different hypotheses, medical hypotheses.

Because of that, I think you cannot have a discussion about these new agents without talking about the biomarkers and the pathology and linking them together with the clinical symptoms. And so I just make three quick points. I think that in science I always find it's most fascinating when we break our assumptions and we break our paradigms, and I think we're about to do that in this field.

So to be specific, we define the disease primarily clinically, currently. Most other diseases that's not true, it's a combination of both pathology and clinical. And to be fair, absolutely fair, we do define the disease pathologically but we haven't talked much about it yet at this meeting, and I think ultimately we're gonna change to doing both. And the biomarkers, in fact, I think do do that. I think low Abeta-42 or PIB positive is indeed doing this. So I think the field is moving in that direction.

I also think that we talk about cognition as the – cognition and functional as a primary endpoint. Again, I think, ultimately, we've got to talk about both pathology and changes in the clinical parameters together. We just had a discussion about that. I think Howard Fillit asked a poignant question, we had interesting discussion on that.

I mean then finally, to bring it back around to this panel, the biomarker pace, I think from a pragmatic perspective, if an agent is intended to modify the pathology and a biomarker tells you that you've done it, it's not a nice thing to have, it's imperative that you have it. And in my mind, if you're gonna alter tau or Abeta or what have you and you don't alter the biomarker, you don't know what you've done in that clinical trial and that's totally independent of the clinical piece.

Let me be very specific here, without naming specific agents, there have been a number of Abeta related agents that were supposed to modify Abeta and they failed in the clinical trials. We have no evidence they modified Abeta.

Jeff Cummings: Yes, David, and then John.

Dave Knopman: I just want to make the point here, I certainly don't disagree with you, but I think for the purposes of our discussion, I think that – and especially before the FDA, that the use of biomarkers for – as an inclusion criteria and the use of a biomarker as an outcome, should be kept conceptual separate.

Dale Schenk: I agree. I agree completely.

Dave Knopman: And I think that we should make sure that we're clear on that. So that's really the key point. The other, if I can just say one thing to John, I think that the data is pretty clear that at least over a year or two, there is a limit – and actually you're welcome to disagree, to what clinical criteria can do in terms of predicting who is gonna decline. If you look long enough, five or six or ten years, you can identify people who are going to decline.

But in mild cognitive impairment, using clinical criteria only, they really have quite a limitation on predicting who is going to decline, and that's why the biomarkers offer a real advantage for enriching the sample.

Jeff Cummings: John, you wanted to comment.

John Morris: So, I still disagree with Dave. I think at the initial assessment, without benefit of longitudinal follow up, one can predict with 92 percent accuracy who will have pathologic Alzheimer's disease, longitudinal, cognitive and clinical decline.

Dave Knopman: Well, but will decline enough to see it in two years in a clinical trial?

John Morris: That's something separate.

Dave Knopman: Oh, no, that's the point. I mean –.

John Morris: Well, nope, nope.

Dave Knopman: I mean this is of a practical perspective.

John Morris: Nope, nope. *(Group laughing.)* But let me come back to biomarkers, because the other aspect of this is that some of these biomarkers may be most helpful in identifying different stages of the Alzheimer process. They don't all necessarily work equally at different stages. For example, low CSF Abeta-42 is most

associated in the pre-symptomatic, the pre-clinical stages. That's when it is correlated with brain atrophy. The lower the CSF Abeta-42, the greater the degree of atrophy in people who are cognitively normal. Once people become symptomatic, even at the MCI or pre-MCI stage, then that relationship changes to CSF elevated tau is associated with brain atrophy.

So there is a shift in when these different CSF measures reflecting brain pathology operate. So, ones in the pre-clinical stage may relate more to Abeta-42, once dementia is expressed and the neuronal degeneration and inflammation and oxidative stress and everything is going on, tau may be a better marker. So I just wanted to bring up that point.

Dave Knopman: And I agree with that.

Jeff Cummings: So, I want to push you just a little bit on that, John, because David has said that he thinks MRI might be the best way of approaching, defining Alzheimer's disease in an MCI population. And I think you just said, maybe you disagreed with that, that there might be other – better ways of looking at it. I think there is an argument that MRI is nonspecific, that it just shows atrophy and there are many causes of atrophy. So, one question would be whether you want a more specific biomarker.

John Morris: I guess I – are you asking me?

Jeff Cummings: I'm asking David but I'm asking you first.

John Morris: So, again, I think it's stage specific. When symptoms are already expressed, however they're labeled, MRI, regional volumetry or stand or whole brain volume may be a good biomarker. But I think the sequence from what we think is the sequence of biomarker events, regardless of clinical status, just in the pathologic process of the disease, the first marker is a reduction in CSF Abeta-42. And that precedes even amyloid imaging positivity. Matter of fact, amyloid imaging positivity probably comes several years later.

Then, as symptoms begin, CSF tau is elevated and brain volume changes are depicted. So I think that's what we would say the sequence of events are.

Jeff Cummings: You would tie your choice of biomarker for entry into the population based on the specifics of the population you were studying.

John Morris: Well, I think CSF Abeta-42 remains low throughout. So it could be a good marker, but I think it's the earliest marker.

Jeff Cummings: Okay, the earliest marker.

John Morris: Yeah.

Jeff Cummings: Okay.

Dale Schenk: Well, sorry, but there's two points about volumetric MRI. I mean there's a pragmatic issue, in that in terms of the statistics and low variability. It happens to be superb that way in terms of sample size and that sort of thing, but that's more of a technical issue. There's also an inherent assumption, which is really a test that we don't really know, and there's a face of validity on volumetric MRI, that you're looking at synaptic density, etc., but we don't really know that, and I think if we knew that for sure it would have even more validity and so there is this unknown that we have.

Dave Knopman: Well, I think you're right, Dale, but I think that that is changing. And again, I don't want to defend this one particular method. This paper is in press in Neurology and it's from ADNI data. You can look at it when it comes out.

John Morris: I took photographs.

Dave Knopman: Well, so – I did really want to keep my job. But the point is that with these techniques, they're in fact – our group has a paper that shows that with this method correlates with Brock staging, for example. And I'm beginning to think, in fact, and I think that the other – that people in ADNI are feeling this same way, that quantitative brain volumetrics is actually the best way to get at synaptic density and neuronal counts better than any other biomarker that we have right now.

Dale Schenk: Yeah, and it's very doable.

Jeff Cummings: To all three of you, we have here a lot of advocacy group representatives, the Alzheimer's Association, the Alzheimer's Foundation, should they have the message for their constituency that lumbar puncture is expected as a part of clinical trials and a part of advancing therapeutics for Alzheimer's disease? John.

John Morris: So, I would say no, unless the study design requires that you are going to have a CSF measure as part of the – either the selection or the outcome process, keeping them separate. But I think lumbar

puncture is – you know, certainly there is a tremendous stigma about it in the community but it can be overcome with education. Lumbar puncture is accepted in other countries without much difficulty. As a matter of fact, it's widely accepted in the United States. People get – thousands of women get lumbar puncture every day, right, when they deliver babies and get an epidural and they have no problem with it whatsoever.

Dave Knopman: They scream for it. *(Audience laughing.)*

John Morris: I mean there are no bad – all of the outcomes that people are concerned about, paralysis and infection and so forth, and people tolerate it just fine. So it's a matter, I think, of education. But in order – my answer, no, it shouldn't be mandatory, we would say that it really should be voluntary but we need to do a better job of explaining its value and demystify or rid ourselves of some of the myths about pain and paralysis and things of that sort, which, done properly, should not exist.

Dale Schenk: Yeah, I sort of have a pragmatic answer to that from the clinical trial perspective. I mean I agree it should definitely not be mandatory, but again, we have to divide up what we had talked about earlier. There is the does somebody have Alzheimer's disease and then there's the business about testing of specific agents.

It's certainly not required to tell you if you have Alzheimer's disease but if you have a beta secretase inhibitor or gama secretase inhibitor, at some point you have to do a study to show that you're altering Abeta in CSF or what have you. So there is a pragmatic issue, as a sponsor that it's almost incumbent upon us to demonstrate changes in some of these biomarkers because of the medical hypothesis for which we are testing.

But I think it's absolutely true, what John was just saying, it's about education and it's not such a horrible thing. And you know, after all the work and all the publications on CSF markers, they do a great job. Not always necessary but for some of our hypotheses we're testing, they're absolutely necessary.

Dave Knopman: Well, I think when John said no, he meant yes. And I will say that we – I didn't promise not to be provocative. That we do want to encourage our advocacy groups to dispel the myth of the malignancy of lumbar punctures, which just simply isn't true.

Where it comes up in my perspective, not just in clinical trials, which has simply not been a barrier in the trials that we've been involved in. We have a population based study in our county, where we're doing – we're collecting CSF in normal individuals, and we managed to have a rate of somewhere over 30 percent of normal volunteers. And this is really critical data.

One of the things about CSF research, especially in the United States is that for many years it didn't include true normals, it was people who were getting other procedures, which are biased. And so to have volunteers participate in spinal fluid studies, people – elderly people is really critical for advancing the research.

Jeff Cummings: Yeah, so I – I mean my message would be to the advocacy groups that I think we should be raising the level of education in the constituency about the importance of lumbar puncture for diagnostic reasons in confusing cases and for therapeutic outcomes in clinical trials. And the larger point that participation in clinical trials is the only way to get from here to better therapeutics is, I think, a critically important message for the advocacy groups to take forward.

Dave Knopman: Make one other point, I think many people might be aware, but John's group was the first to show that the CSF Abeta reflects the PIB scan extremely well, and as John pointed out, it actually anticipates PIB positivity. So one of the insights that PIB scanning and the CSF work has shown that it validates the CSF. It provides a validation of the CSF findings for Abeta. It's much less expensive. It's much more accessible to a broad number of investigators and patients. And so that, I think, has also been something that in the last few years represents a real advance.

Jeff Cummings: And Dale, I'm gonna put you on the spot now. So, John Morris says he can make the diagnosis of very, very mild AD with 92 percent accuracy. David, unfortunately, has admitted that he just can't make the diagnosis. So in a widespread – in a big, multi-center clinical trial, how competent do you feel about making the diagnosis clinically without a biomarker?

Dale Schenk: In terms of multi-center, not very confident, honestly. I think the inclusion of a biomarker for identifying those who are either very early or very, very mild is gonna be required, at two levels. Not just about identifying the folks, but also having a rigorous standard or a baseline for how to select them that can be even broader.

Because, you know, again, at this meeting and many meetings like this, we always focus on the clinical trials and the approval process, but once an agent is out in the field for physicians to use a lot of things happen and then we're gonna get to even the next level of knowledge, so to speak. And if a biomarker can be used to identify patients, I think that's great.

I'm all for dividing up patients to identify them better in terms of who – it's appropriate to have subsets that respond to therapies if that's the way it works.

Jeff Cummings: John, you had a follow-up.

John Morris: Well, so we're talking about making the diagnosis earlier and earlier in the symptomatic course, and we all agree that that is important. We may differ a little bit about how to parse people out and identify them and what to call them, but I think we all agree that that's important.

We have to understand that – and I know everyone does, but just to make it quite clear, we're talking among ourselves as people who are doing this for research purposes as well as clinical purposes in specialized centers, have to recognize that half of the people or more in the United States who have dementia are unrecognized by their primary care physicians.

So it's not, you know, we can talk about cognitive testing and so forth, the fact is, if we get an agent or agents that are absolutely helpful in treating people, we have to somehow empower practicing physicians to better diagnose Alzheimer's disease. I think there are lots of reasons why they don't, one of which is they don't see much value in the currently approved medications, that's one reason. So if we had an incentive, you know, we have an actual drug or a group of drugs that do help the disease quite effectively, I think that's one aspect.

But I also think that, you know, practicing physicians are poorly reimbursed, they have little time, they lack knowledge, and I think biomarkers would be a tremendous advance for improving the recognition of the vast majority of patients who cannot find a physician to diagnose or manage them.

Jeff Cummings: Dr. Katz, I wonder if I could ask you to comment specifically on the involvement of non-demented Alzheimer's disease patients in clinical trials and on the use of biomarkers as a way forward to identify such a population.

Russell Katz: Yeah, I think – we certainly applaud any efforts to identify patients earlier and earlier in the disease. So again, we tend to – when it comes to something like making the diagnosis of Alzheimer's disease in the absence of clinical symptoms or particularly in the absence of dementia, we tend to take our cue from the expert community about whether or not we have new diagnostic criteria that allow us to reliably identify these people. So we talk about the Dubois Criteria or whether there is other criteria.

When there is s consensus in the field that, in fact, the Dubois Criteria identified people who have Alzheimer's disease and therefore are expected to have Alzheimer's dementia down the road, then I think that's an appropriate thing, that's a thing we should applaud.

Of course, my fear, of course, if that's what it is, is that we'll get so good at diagnosing patients 15 years before they are gonna become symptomatic that the use of biomarkers are primary outcome measures will be endorsed in the absence of real good information that they actually tell us anything. So but – you know, in my list of fears, that a good one to have, you know, so I don't mind it that much, if we can get to that point.

But no, I think the attempt to diagnose patients earlier and earlier is to be applauded. Again, how will we assess them, how we assess drug effects in those patients is, I think, largely what we're talking about here, it's not so simple, necessarily, but the earlier the better.

Jeff Cummings: And when biomarkers are used to define an inclusion of population, how will that affect labeling?

Russell Katz: Yeah, obviously, as you said before, that's a question that a lot of people have, it's a question that we have. You know, I think as we get sort of smarter and smarter about these things and have a way to better identify who people are who are potential responders to a treatment, I think a real argument could be made that those tests ought to be used to identify those people or it should be described in the labeling or perhaps even the indication should be limited to those people.

This is happening now on many fronts with genomics and as we get smarter about saying, well, these people have this particular genetic profile and those are the people who respond to our treatment, why would you not want to make that information prominent in labeling.

So, I'd like to hear what the arguments would be for not having to list in labeling who these patients are who have been studied and therefore, who we expect to respond. Every drug is potentially dangerous and if we can exclude 30 percent of people or whatever the number is that we would've included on purely clinical grounds, if we can get that group down to the people who really do have the potential to respond to the treatment and not just be hurt by the treatment, why wouldn't we want to do that?

Jeff Cummings: Here is one possible scenario. A large number of Alzheimer patients live in rural areas where lumbar puncture, for example, is going to be maybe less accessible, the availability of specialists, neurologists is not so available, but payment might be limited by the labeling to just those people who have the biomarker.

So now I'm in Wolf Point, Montana and the nearest neurologist is in Billings and I'm gonna be denied payment for my treatment for my Alzheimer's disease, which in this case, my private practitioner has recognized. That would be a fear I would have as a clinician, that there might be limitations such as those.

John Morris: But there's also a neurologist in Bozeman. *(Group laughs.)*

Russell Katz: Well, I guess it depends – and was said before, it depends on how huge the – well, it might at least, one factor might be how large the discrepancy is between how well you can make the diagnosis on clinical grounds alone and how well you can make the – how better refined your diagnosis is when you use the diagnostic criteria. If you can get 92 percent, for example, depending upon the stage of disease, just on clinical basis alone, then there's an argument to be made that you don't have to do more sophisticated tests to identify these people.

But again, I think as we get smarter and more sophisticated about who these people are and what their particular biochemical abnormality is and are they likely to respond to a particular treatment, I think it would be important information to at least put in labeling.

Now, we do something like this now. We say a drug is approved for mild to moderate Alzheimer's disease and then we describe in the clinical trial section of labeling what tests were used to identify people as having, I guess, both Alzheimer's disease and also mild to moderate.

We don't put in the indications section, you have to a score of X on the MMSE in order to get this drug, but we do describe in labeling what the rules were that got people into trials. Although, I understand that that poses some problem for reimbursement as well, or potentially does.

So I think it depends on a lot of factors. I do think, though, at least you would describe somewhere in labeling how you identified these people. But again, I think if we get to the point, particularly earlier and earlier where the clinical criteria alone might be less reliable, we have to seriously consider the fact that if we're gonna be throwing out 30, 40 percent of people who on clinical grounds might have MCI but who don't have Alzheimer's disease, I think we have to seriously think about making that a prominent part of labeling.

Jeff Cummings: Thank you.

Dave Knopman: I would just – it sounds much more appealing to hear you say that it's in the labeling as opposed to it's a requirement for receiving the drug that the tests be done. That does, more likely, throw it back on the insurer and perhaps that would be an area that would be more negotiable and –.

Russell Katz: Well, right, but as I say, it does depend upon, I think, what we think the discrepancy is between being able to reliably diagnose on clinical grounds as compared to when we use these additional more sophisticated tests. There certainly are examples where – or we're contemplating examples where you'd have to have a particular – this is particularly on the base of sort of genetic testing, you have to have a particular test done at the time and you have to be determined to have this particular genetic defect in order to get this particular treatment.

So there are cases where there is a very tight linkage between a test result and the indication for the drug. But again, depending upon these other factors, it might just be that it would be described – what would be described in labeling is how these people were identified and not necessarily make it a requirement that the test be done.

Dale Schenk: I was just gonna comment too, I mean again, for the audience. You know, if somebody shows up with dementia and they have diffuse Lewy Body disease or frontal temporal lobe dementia and you know, one was synuclein related and another one is tau related and the physician is thinking about giving them an Abeta

approach. It certainly won't help them. Now, I think the good news there is they're not as often seen and are distinguished colleagues here would not misdiagnose them. I think out in the field it does get misdiagnosed frequently and you would hate to give a complex drug to somebody for which they have no hope of benefiting.

Russell Katz: Well, right, and if you may recall, we had an advisory committee, it's now, I guess, a number of years ago, about what ought the standards to be for approving a drug to treat MCI, and one of the requirements that the committee recommend was that there be a – at least one trial in such a package, in which it was demonstrated that sort of the average practitioner could make the diagnosis reliably. That was an important part of – we weren't talking about biomarkers at the time, but that was an important part of the discussion at the time.

Dave Knopman: I just wanted to just make the point just of this data, which isn't – we can do better, but just if you look at the ApoE4s and the ApoE4 negatives, what a difference there was in terms of the outcome and perhaps the people who did not carry the E4 allele, there was some decline but there wasn't very much. It was much more in the E4s. And I think that this may be the least effective of the biomarkers.

And if we could identify people who don't progress, I think that this is a win-win situation, that we've identified people by biomarkers who have cognitive impairment that appears to be non-Alzheimer's, or non-progressive. We might not be able to distinguish those, certainly at the primary care level, and treat only those who we expect to progress, just like we wouldn't want to treat chest pain people who had normal EKGs and normal enzymes and that sort of thing.

Jeff Cummings: Okay, I'd like to open it up a little bit to the audience just on this question of mild cognitive impairment. Thank you, Dr. Katz. In participating in clinical trials, because then we're going to go to the open discussion that will be led by Dan Perry. So are there any questions on this – or comments on this specific aspect?

Louis Kirby: You know, as usual the microphone stand is too short. Louis Kirby, I'm with the Critical Path Institute and the Coalition Against Major Diseases. Short background, our task is to identify just exactly what this topic is, and put a package in front of the FDA for consideration of it in use in clinical trials.

The first question really has to do with, has any data been used to verify the Dubois Criteria, other than what was in the article? Has the ADNI data, for example, been used to help validate this criteria?

Jeff Cummings: There are several studies of the Dubois Criteria. They are primarily retrospective so far, and I don't think it's been published on the ADNI dataset yet.

John Morris: Not that I know.

Jeff Cummings: Yeah, so –.

Dave Knopman: ADNI does not include the free and queued selective reminding, which is a key – test that Bruno focused on, which was John's appropriate objection. But in terms of the biomarkers, the data appeared in the sense that the Shaw paper, the CSF conversion, there are certainly repeated, not just in ADNI but also about hippocampal atrophy, the PIB data was a part of the Dubois Criteria but serves the same role. So I think in so many words there is validation.

If I can just make one comment about the free and queued selective reminding, which I think is – it is, I think objectionable to pick one memory test, one specific one. There is some data emerging on that. It turns out that it's fairly sensitive but it is terribly nonspecific. It identifies, as John was saying, many people who don't have a cognitive disorder, who still score in the abnormal range, that's just one of the problems of memory tests in older patients.

Louis Kirby: So do you have one or two favorites that you would substitute for the free and queued selective reminding that would say these guys have some level of cognitive impairment. It's an amnestic type of impairment, some form of delayed recall, say?

Dave Knopman: Well, I don't think the different places have different tests that they like. The combination of logical memory and AVLT or California Verbal Learning. I think – but it's a whole package, it's not just a test. It's the clinical sense that there's something that has changed. And I think focusing simply on a quantitative score based on questionable normative data is a problem.

Jeff Cummings: Let me defend the Dubois Criteria just slightly. I feel France being attacked here. The Dubois Criteria require that there be a history of progression of cognitive decline in addition to failure on the

mental status test. So there is an attempt to put it in an historical context. And while the free and queued recall is championed as the best measure of hippocampal dysfunction, which – and I think that's up for grabs, I think it's not actually required as the only way to define patients who have the critical type of explicit memory defect that he is looking – that we are looking for.

Dale Schenk: Just a quick comment too that partially answers your question, though indirectly. A number of the different pharmas and biotechs, the Dubois Criteria has been valuable in the context it has opened up a discussion about how you would select, who you would select, looking at the ADNI data very carefully to identify those who will go on to progress, etc.

And although nobody is officially talking about it, in the hallway there's all types of discussions by many of the pharmaceutical companies to consider doing exactly that. In other words, identifying patients who meet some of the biomarker criteria as well as some of the cognitive criteria. And the reality of all these things is always pragmatic, somebody's gonna do it. If an agent works and the selection criteria was successful and there's correlations on the biomarkers, it'll slowly be discussed by advisory boards and utilized more and more, I think.

Jeff Cummings: One more question and then I'm going to close this session and we may have more questions for the open session.

Louis Kirby: Well, yeah, bank on it, but one of the other issues that comes up in terms of trying to get standardization, is if we have a set criteria for the designation of a likelihood to progress to Alzheimer's dementia in the inclusion criteria on the protocol, you may have some issues with which lab do you pick, which protocol for volumetric assessments do you use. Is there some consensus on how you would select the lab for your CSF marker, analysis, the protocol for analyzing the volumetric imaging? I mean these issues are gonna come up, particularly when we're looking at trying to validate these from one dataset to another. So I'd like to hear your comments about standardization of these applications.

John Morris: So I think that's one of the great attributes of ADNI, at least in the research community, it is setting the standardization of fluid biomarkers, imaging biomarkers, and it's –. You know, things will change but it, at least, is getting all of the centers involved in ADNI to do things the same way and other programs outside of ADNI are often using the ADNI protocol. So I think it has done a great job.

The problem that isn't standardized was the first part of your equation, that is progression to Alzheimer dementia. You know, there is no standardization of when a clinician determines who is somebody – that somebody is now demented, that varies quite widely. But we all talk about it as if we know what we mean among each other. But in fact, it can vary quite widely depending on your threshold and the quality of the information you get about functional impairment. So that's where we need standardization.

Jeff Cummings: And there have been some studies showing that logically it's better to do all the CSF in one batch rather than in serial batches, but in ADNI, of course, it's done in serial batches, because you don't even know when the end of the study is going to be. So there are some choices, practical choices made for ADNI that one would actually not choose in a clinical trial where you knew that you were gonna have an LP at 18 months and an LP at baseline and you want to run those at the same time to see if there's been a change.

Neil Buckholtz: That's not exactly true. I mean – just to point out that what is being done in ADNI is that it's being – the CSF is analyzed at certain points when there are the one-year data, the two-year data. But actually what Les Shaw is doing is going back and analyzing – he's doing the two-year data, going back and reanalyzing the one-year data and finding out how to equate what was found at one year with the two-year.

So, again, in terms of this kind of standardization, I do think that it will provide a way of looking at serial samples and being able to equate what was found at an earlier analysis with what was found at a later analysis. So – and Les has mentioned that in the recent paper.

But I think that really is an advantage of trying to look at longitudinal data where you're trying to not wait till the end of the whole study to analyze all of the data. And again, I think it's an advantage, just to make a point, as I was mentioning to somebody in terms of even the imaging, the issue of the standardization of ADNI in terms of quality control on quality assessment, what had previously been done in a lot of the industry trials that included imaging was basically to wait till the end of the study to look at all the imaging data. And if there were a problem you wouldn't know it until the end.

The advantage of ADNI in terms of the standardization is to look at the data as the data are coming in, so you have a chance to go back and rescan somebody if that's a problem.

So, I think as you mentioned, all of these kinds of standardization, the quality control, the quality assurance is really critical in terms of using biomarkers in clinical trials. And I think that this is what has to be appreciated as we go forward with the biomarkers, that there are ways of doing this, it's much more difficult but it's critical if you want to use biomarkers in the context of clinical trials.

Jeff Cummings: Yes, thank you Neil, and that's a terrific point, and I think ADNI has defined the standard for multi-center collaboration for biomarker research.

John Morris: May I make just one very small point. In terms of standardization, it really is critical, just with CSF, for example, depending on the time of day the lumbar puncture is taken, the CSF level of Abeta-42 or tau may vary quite a bit. So if you're doing a multi-center study, everyone should standardize to take the CSF at the same time of day so you can compare the results.

Jeff Cummings: Great. So I'd like to thank the panel. *(Applause.)* And I would like to invite Dan Perry to come up who will run the final open session of the meeting. Dan.

Dan Perry: Actually, it would be best if you could all stay in your seats. This may be the subject of some of the questions. And because we now want to give everyone in the audience that has not had a chance to engage with our clinicians and with our scientists a chance to do that. This is sort of the bonus hour, as we move toward a close.

So I'd like to ask Dr. Black, Dr. Siemers, Dr. Rockwood, if you'd join us up here as well, and Dr. Fillit, as well. So, if we have anyone left that can ask an intelligent question, and I know that we do.

Let me say, my name is Dan Perry. I'm the Executive of the not-for-profit Alliance for Aging Research, and I am privileged to be the chair of the ACT-AD coalition. One of the three co-hosts of this. And let me just say how grateful I am to all of you and all of our experts for allowing patients and research advocates like ACT-AD, like the Alzheimer's Association and LEAD, groups that also embrace a lot of significant science, but essentially we are people that are coming at this from nonscientific backgrounds, we're not part of industry, we're trying to come in from outside. But we're

the people with the most at stake, and so it is wonderful to allow the patients and the advocates, who, after all are laypeople, by and large, to be at this critical juncture an to encourage this kind of, what I think this morning, has been an extraordinarily rich, honest and flexible exchange of ideas.

I'll say a bit more in closing, but for now, let's open things up and take any questions of any of those that you've heard this morning. Gail, in the back of the room, and we're gonna pass this around.

Gail Hunt: Okay, thanks. Gail Hunt, National Alliance for Caregiving. We've talked quite a bit about the primary care doc as the first person who does the diagnosis of the person with – or even maybe recognizes the person with Alzheimer's, but actually it's the family member that usually recognizes the symptoms and then brings the person, brings the family member to the primary care doc to get some kind of testing. And has been alluded to, not every primary care doc is really enthusiastic about doing a diagnosis. So it may be then that the family member has to – if they're lucky, they can get a referral and six months later they can actually get a diagnosis from a neurologist.

I'd really like to hear a little bit more about how – what's the role of the family caregiver, perhaps in having a better awareness of Alzheimer's symptoms so that –. And then, where, really is the best place for them when they begin to recognize something's wrong with dad, where is the best place for them to go, where they're gonna really get a true diagnosis?

John Morris: Dan, may I take a stab?

Dan Perry: Yes, please.

John Morris: So, very cogent comments. I think I'll answer the last part first. The best place to get a diagnosis is Mayo Clinic, Rochester. *(Group laughing.)* Kidding aside. You're absolutely correct. I think, you know, Jeff and Ken and Dave may have the same experience. It rarely is the individual affected by early symptomatic Alzheimer's disease who generates the initial visit. It's almost always the family. So the family recognizes the change.

You asked, how can we do a better job of informing the family to recognize those early signs. I think that's a huge job. I think the Alzheimer's Association, nationally, has tried to take that on, Ten Warnings Signs of Alzheimer's Disease, but it is a public awareness campaign, and I think that we should empower the

families to recognize and seek attention when there has been – as I've been saying all along, a change in that person's cognitive ability to carry out their usual activities, they're slipping compared to where they were. That is not normal if it's interfering with their daily function, so that does deserve evaluation.

I talked about practicing physicians. I want to be very clear, there is nothing that is essential about a neurologist or a geriatric psychiatrist or a geriatrician, any physician who is interested in Alzheimer's and who can take the time to get the history and to do the evaluation is equally as good as me or anyone else. So it doesn't have to be a neurologist. By the way, there are way too few neurologists to diagnose everybody in the United States who has Alzheimer's. We have to get education out to primary care physicians, who can do a great job if they take the time and their interested.

But I think the – I said this earlier, I think the big thing on the medical side, there are issues of how much time we have, how much reimbursement people get. I think the big thing will be if physicians believe that it is important to make the diagnosis because we can do something about Alzheimer's. I think that's the incentive. And so that's why we're all engaged, passionately engaged in trying to develop truly effective medications. I hope all physicians would do this without that incentive but I think that's what it's gonna take.

Dan Perry: And Dr. Fillit is a geriatrician and sees a lot of patients and families, and I think you'd like to say something on this as well.

Howard Fillit: Yeah. I'm an internist and so a lot of primary care. And I think that contrary to some of the myths about whether doctors are taking care, I mean I think doctors are not paying enough attention to the disease. But on the other hand, we know that 70 percent of the prescriptions written for Alzheimer's disease are written by primary care doctors, and it is a primary care illness.

I think that along the lines of what John said, that in the medical model, doctors like to write prescriptions and in a short office visit with an average length of time of about eight minutes in the United States, there's not a lot of time for counseling. So we do need to help doctors figure out a way to get counseling done. And you know, that can often be done, as John alluded to, through ancillary providers such as the Alzheimer's Association's Caregiver programs.

But I disagree with John on this idea that we need treatments to take care of these patients effectively, and I'm sure John agrees with this, that all of us have been caring for Alzheimer patients for many years, and for many years prior to the time when there were any drugs on the market. And I know that I like to say that I've never met a family or patient with Alzheimer's disease that I couldn't help. And I think for all chronic diseases, we don't have cures for almost any chronic disease that I can think of, and so we're always combining treatments, whatever they are, with care management type of approaches, that help patients to cope with chronic diseases that are ultimately progressive.

The only other comment I wanted to make was about the spinal tap, and I know that those of us who have been around for a while remember the days when spinal taps were actually part of the differential diagnosis of Alzheimer's disease and we had to rule out neuro syphilis. And so if you looked at the clinical practice guidelines or their equivalent in 1980, a spinal tap was kind of part of the differential diagnosis.

And I know as an internist that when I trained, we did spinal taps all the time, in the emergency room for meningitis or for head trauma to – you know, this kind of thing, rule out strokes and just as a primary care. So I think the idea that the spinal tap has to be done by the neurologist is also a little bit overdone. I think that internists and primary care doctors can do spinal taps and that if we get an effective drug that requires a spinal tap, let's say, for diagnosis or monitoring, that that could also be done in the primary care setting.

Dan Perry: Great. Thanks, Howard. Anyone else on the panel want to respond to that question?

Jeff Cummings: Dan, maybe just one more comment, Gail, is that as we move towards earlier patients, which is what we're talking about today, there is more self-recognition of patients, that they're somehow failing in their work. They don't always think, well, gee, I have Alzheimer's disease but they – gee, work is getting harder for me. There's something wrong. And we're certainly seeing patients who are able to articulate their mental state better as they have earlier and earlier disease.

Dan Perry: Another question.

Meryl Comer: This question is really to try to shift the paradigm. I've been a caregiver for 15 years and I have seen early onset and I have seen

late stage with my mother. So, now at home I have two. So, I am not a scientist, I'm a clinical observer.

I want to shift the paradigm to what I call the worried well, and if this is the dark side of longevity that we all have to worry about, what is the added value to the experts in front of us of a population that went online and tracked themselves until there were signals that they could not ignore and then came to centers of excellence? What is the value-added of a 60-year-old who may have one marker because they've done the genomic testing and we offer you 10,000 of us and we say what is the value of tracking us? Will that speed up the process to find a disease modifying drug.

Dan Perry: Meryl Comer of the Geoffrey Beene Foundation. Meryl, you need to let us know about you.

Dave Knopman: Well, let's see. Well, the first issue, Meryl, I think that the – I think what you were saying is one of the potential downsides of the increased public recognition of Alzheimer's disease is that many more people are overly concerned about their memory and should they have an outlet, is that what you were saying?

Meryl Comer: No, I was trying to see if there was any scientific value in tracking the populations. We're trying to move earlier and earlier and engage people –.

Dave Knopman: Right, well –.

Meryl Comer: I'm much more likely to have the lumbar puncture if I have been engaged earlier in the process –.

Dave Knopman: Right.

Meryl Comer: Observed symptomatically, I have some data that may be of value to someone that may be-and used. I mean you have online patient communities.

Dave Knopman: Well, what –.

Meryl Comer: Who are doing this already.

Dave Knopman: Right. I think from an epidemiologic and a methodologic perspective, doing it on a – at least up to this point, doing it on a geographically-based population of cognitively normal people definitely needs to be done. As I was saying earlier, in Olmstead County, Minnesota, we've recruited almost 3,000 people between

the ages of 70 and 89, initially who were nondemented and we are following them on a roughly yearly basis.

The reason that I say geographically defined is that you don't want to preselect just those people who are motivated to sign up, because that introduces bias. But there are ways of collecting data on normal volunteers who are interested in an unbiased way and those kind of studies are going on and they are really critical.

Because actually what you said is true, we've been talking here about people who have the earliest clinical manifestations of the disease. What we really want to find out about and this is something John has been working on for many years as well, is what's going on in the people who are cognitively normal, ten years before they're destined to become impaired. How can we recognize that, how can we learn about the biology, and that's really critical. But you've got to do it in an unbiased way. It takes big numbers but it is definitely doable and there are a number of studies that are working on that.

That was the first part and I forgot the second part.

John Morris: Well, just to augment what Dave says, all of which is correct, and I think that Dave's caution against bias is correct, but we took the approach when we wanted to study the sequence of biomarker changes in cognitively normal individuals throughout the lifespan. So we now are looking at people as young as 18 but we began at age 45 and up, who are cognitively normal and we asked them to do everything, not just testing every year and – but lumbar puncture and MRI and PET and genetics and everything.

And we said who would volunteer for this very difficult study. The people who were most interested, there's biased, are the children of a parent who had Alzheimer's disease. So we call it the Adult Children Study and we were overwhelmed with volunteers. We had too many because we also wanted a group of individuals whose parents didn't have Alzheimer's disease to be the control group and follow them along. But it's enormously – just as you say, how can people track themselves over time or learn about the disease and the people who are most highly motivated, I think, are the children of Alzheimer patients.

Dale Schenk: And there's just one comment too, on sort of a technical and pragmatic level. You know, any type of diagnostic test can either be sensitive, specific or both, and so there's a number of tests out there or computer programs you can use that can sort of tell you

how you're doing with your memory. These tests, by definition, can be very sensitive, but they're very, very nonspecific. So they have a strength and a weakness, and I won't speak for the rest of my colleagues, but I've always felt that at some point in the future, they will have a very definitive place in terms of just a general exam, to at least tell the general care practitioner how that individual is doing. And if they start declining, that will invoke a number of other tests to pull back in the specificity of that decline, which could be for a lot of reasons, totally unrelated to Alzheimer's. It could be depression, for example, it could be a lot of different things. You have to be very, very, very careful with it, but it's early days on this.

Dave Knopman: If I could comment on the genetic issue, the second question. In fact, there was a study of people who had the E4 allele that was just published in the New England Journal last week by our colleagues in Arizona and that's really fascinating work. These guys enrolled people who were volunteers in their – I can't remember how young they were, 40s or 50s, and have followed them for a number of years and were actually able to detect a change in cognition in these people who were still cognitively normal amongst the people who had the E4 allele.

So these kind of studies are very informative. They give us a – it's a critical piece of information for telling us, as John and others have mentioned, that the illness has its seed, so to speak, at least a decade or probably more, well more, before people developed the earliest symptoms. I mean and ultimately, maybe not in this decade or the next decade, but some time, that's where we would really like to target therapies for.

John Morris: May I make a – Dan, may I just make a comment, it's really not a – a non sequitur in a way, but when we're talking about biomarkers, we often talk about molecular biomarkers and I think they're very important, but there are many other biomarkers, we mentioned cognition. One thing – and this gets a little bit at what Jeff was talking about earlier, that we don't track but may be very informative, is personality.

And Bob Wilson, at the Rush Group in Chicago has been a leader in trying to determine in people who ultimately develop Alzheimer's disease when are there personality changes and when they might occur. And we've found, again, the same sorts of things, that even in mid-life there is a distinctive personality profile that does seem to be associated with later development of Alzheimer's disease. And among those personality traits are

inflexibility, so more rigidity, self-centeredness and something called neuroticism, and you know, interest that goes along with self-centeredness.

And at our place, when our analyses were done they came to me and said, John, that's you, so –.

Dan Perry: Okay, well, we'll keep an eye on you. Doctor, heal thy self.

Dan Perry: Dr. Rockwood.

Ken Rockwood: So I want to speak to the point of the merit of people being able to track themselves online and share observations. So you've heard a certain skepticism, right, there's bound to be a lot of chaff with that, but I think there would be some very good wheat too. And the thing about the ability to aggregate the information that comes online is that it's reasonably straightforward. There are reasonably straightforward ways to look for important signals. I think that is worth a shot.

I can share my own experience with the website we have, where people go on and build a symptomatic profile. And we've got data now on about 500 people who have gone on in the absence of a diagnosis. In about 300 of them, it's clear that the diagnosis is Alzheimer's and they just haven't been able to persuade anyone of the narrative of that yet.

We've gotten some really interesting observations that have come out of letting people say what they want online without constraining them. So, for example, I was not aware till I learned about this, and then I've tried it out in my clinic and it seems to be true, that an important proportion of people with early Alzheimer's have the problem of shadowing. They can't let their spouse out of their site, and that's much more common than I had thought. It's often very deeply shameful to the spouse. They will actually follow them into the bathroom. Like, they can't bear not to have them in view the whole time. Where does that come from? I'm not sure but I know that it's a symptom we need to look into more now that I've been in the field for a long time and certainly had not come across that at all to the extent that I've seen it now, once I've been prompted online or by what people have put in online to look for it.

So, you know, I would think it's worth a shot and maybe some really important things to be had from that. It would need to be a fairly automated process and something that is set up in a way that

allows for reasonably straightforward signal detection to be done. But I think the web is an entirely underexploited resource in trying to define the variance in an illness which is as highly dimensional and as common as this is. So I'm enthusiastic for that.

Dan Perry: Ron.

Ron Black: Well, I'll maybe try to take a different tack in answering that question. I think that to a large – with the disclaimer that I work for a drug company and it's been over ten years since I've seen an Alzheimer patient in a clinic. But basically, I think that ultimately, it's going to be the opportunities to intervene that are going to drive the earlier and earlier diagnosis of Alzheimer's disease.

And when intervention will make a difference, then it's going to become suddenly important outside of the academic community to diagnose Alzheimer's disease early. And in fact, it's going to have to go to primary care. Now, this may seem impossible because of the difficulties of diagnosing Alzheimer's disease that – and how the sophisticated sort of thinking that we've seen here, but look again at the oncology model. In fact, when it became necessary to diagnose, say breast cancer early, a whole infrastructure evolved. Now mammograms – and this is very complicated medicine, but in fact, with primary care physicians at the frontline and then a second line and a third line, in fact, now, we're doing a much better job of early intervention in cancer.

And this is the model that's gonna have to be driven by primary care physicians and by the opportunity for a meaningful intervention, which I don't think we have now. I mean we have treatments now, but in fact, when you start the treatments is not gonna make a difference in terms of the long-term outcome.

Dan Perry: Dr. Siemers.

Eric Siemers: Yeah, so this is along the same lines, and let me just follow up on that. I think one of the things, when you get involved in these discussions, it's really interesting to see how quickly people start to gravitate to start to talk about what's really pre-symptomatic, screening and diagnosis, if you want to call it that. Now, to get back to John's opening remarks, I think we were going to try to focus on early, but not really pre-symptomatic. But the fact that it comes up, I think says that conceptually, people understand that the pathology starts a long time before you develop symptoms.

But to get back to sort of early symptomatic stages, I did want to follow up on a related point from Dave's slides. He actually covered a lot of good points in just a few slides, and one of the things that came out of his discussion is whether we can enrich trials for patients who we think will decline more quickly. And I think that's a thought and it's gets a little bit into the semantics, but it's a thought that deserves a moment.

At least in my mind, when I think about enriching trials, I think that's something that's fine for phase two, where we're just trying to figure out whether or not you can get a signal from this treatment that you're testing. But in terms of a phase three trial, then you get into a lot of issues about how generally applicable is your trial and the results of that trial. So in an ideal world, I think what we'd want to do is in a phase three trial, we would want to use generally established diagnostic measures as our inclusion/exclusion criteria.

Now, what's interesting and actually promising about Alzheimer's right now is that we have these two parallel processes going on. One is improved diagnostics and the other is those of us who are working on therapeutics. And so, sometimes the therapeutics world kinda pushes the diagnostic world a little bit. But I think what – and we may not be able to do this perfectly, but what I would like to be able to see is in future phase three trials of these very early subjects, that we use as inclusion/exclusion criteria, something like the Dubois Criteria or whatever it turns into.

There is a movement afoot, actually at NIA and Tony Phelps may want to comment on this, about changing the diagnostic criteria to incorporate those biomarkers. But we really would want to do that in a way that it's generally applicable. United States is probably a little bit different than either other parts of the world and not every technology is going to be available in every part of the world, but as we start to think about treating people earlier and improving our diagnostic process, we need to think about how that could be done from a public health standpoint.

Dale Schenk: Sorry, I hate to comment because I already did, but Eric, I have to disagree with you. And just to show that at the industry level, we don't all agree either, but –. I agree with most of what you said, but I actually think – if you look at the neuropathology of any two Alzheimer's patients, they're never the same, ever.

And – they're similar and there's a lot of similarities, but I think at the end of the day we're not gonna be treating Alzheimer's disease. I think we're gonna be treating those with plaques and then we're

also gonna be treating the tangles and we're gonna be treating the cardiovascular aspect. We're gonna treat the different components and they will have impact on the clinical picture, and I think perhaps one of the reasons the trials have failed to date is that we do try and set general standards, but remember, it's a circular argument. As long as we assume Alzheimer's disease is one entity then it is one entity.

And I'll just put out there that it – I'm not gonna say it's ten different diseases and go that far, but I will say that I think patients differ and particularly at the early phase. And so that's just a general comment, and I think different treatments, depending on their target, will focus on different aspects of the disease and pathologies, and we will have to identify those that will benefit from a given therapeutic.

Dan Perry: Anyone else on the panel on that question? See what you uncorked, Merrill. Thank you.

Meryl Comer: Let me just add in response to John's observation about personality changes. The study that needs to be done is through the divorce courts, because the line – and we have a bad sense of humor in the community of caregivers, is that the line is Alzheimer's saved our marriage, and what it really says is those behaviors, John, that you described, are really the triggers for separations, divorces, and you find a loved one is sick and you come back and then are lost with them in the disease.

John Morris: That's very interesting.

Dan Perry: John Dwyer.

John Dwyer: If I may have your indulgence for one more question.

Dan Perry: Absolutely, you do.

John Dwyer: I'm rigid and narcissistic about a concern that I have if we roll forward, and that is as you guys move forward and soon have therapies, this issues of biomarkers is gonna come back and potentially put us in a box canyon, if not anticipated properly from a reimbursement point of view and that's gonna express itself in two different ways. And Howard and I have had this conversation before.

Dr. Katz, you guys got to write labels the way you see clear, but however you write the label will be the default starting point for

reimbursement when you go to comparative effectiveness over at Medicare. You know, that's not your responsibility but the rest of us have to worry about if there's a label that says effective only with ApoE4 or a biomarker of a spinal fluid test that shows certain criteria, that's gonna be a threshold issue that comparative effectiveness is gonna say, we're gonna reimburse for this or perhaps even a higher hurdle rate.

But you have another issue and that is in the issue of just pure diagnostics. The research community – this is the rigid part, John, the research community is, because of the necessity to move forward, is failing to recognize that some of the biomarkers you're talking about and the diagnostic procedures simply are likely not to be reimbursed by Medicare as a clinical diagnostic test. They're not gonna pay for an MRI every time somebody needs to figure out whether they can or can't take a drug, even if you guys approved it as a diagnosis to get a drug.

And with all due respect, John Trojanowski is not here to protect himself, but John and I had a colloquy not long ago about the lumbar test. And my wife told me, and she is my marketing study of one, lumbar test – and she had a wonderful experience on her C-section, she said, those people swim in 55-degree water nude in Scandinavia. John said, "In Scandinavia everybody has a lumbar test and they think that's the greatest thing in the world. They have no problem with it. They should do that in America." She said, "They swim nude in 55-degree water, live in the dark six months a year and drink vodka and eat herring. Now, that's not the United States."

And so as a marketing guy and as a businessman, we can't look at trying to offer to the testing community at the primary care physician office, a lumbar test as the clinical differential for whether you're gonna be able to prescribe the drug or not, because I think there's gonna be, especially in the pre-symptomatic community, a real issue about whether they're gonna go and get nailed in the spine and never be able to play tennis again –. Men are cowards, by the way, women might do this but men won't.

And so as a consequence, I really worry that the research community is not anticipating these issues of biomarkers in the diagnostics as you're bringing forward your clinical studies. And if we're not looking at biomarkers that give us that grounding, we're in real trouble.

And I think the for-profit community would love to come to the FDA with diagnostic tests that would track you guys and not have too many false positives and not have too many false negatives, but it's got to be at a price point where, you know, 5 million patients and 15 million worried well can get the test and figure this out. That's a big number that government's got to pay for, because Medicare's the likely reim-. So with that, I open it up to comment, but this needs to be anticipated.

Dave Knopman: I'd like to know what you have against herring? *(Group laughing.)*

John Dwyer: It's great.

Dave Knopman: Well, I think actually – I think we do need to test out those models, but you know, there is another side of that, John. If the insurers would know that fewer patients were going to be treated, including patients who didn't need to be treated, perhaps they would be induced to pay for those tests. Now, maybe that's just naïve on my part, but it seems to me that one possibility here is if we have greater specificity of who we treat, that that would be financially – that would be a financial incentive for those tests to be reimbursed.

I'd also just like to comment, and I think Howard mentioned and Ron, that as we – in other diseases, where tests have been shown to be of benefit in identifying who is at risk, the field in those areas have moved there. And so what we need to do is if we can show once we have an effective drug that the biomarker identifies those people who benefit and the absence of the biomarker abnormality identifies people who don't benefit and are at risk, that's a real plus, I think.

Dale Schenk: It's cost effective.

Dan Perry: John, you had a comment.

John Morris: I'm going to get this undoubtedly wrong because I'm ignorant about what actually happened, but I think, John, the way we in the research community are thinking is let's start with the best defined – best characterized group including the use of biomarkers to help us with our characterization, see if the current candidate drugs are effective and then broaden it out to the larger community.

And here is where I'm ignorant. I think this is what happened – Rusty, you may know much better than I do, or others here, I think this is what happened with statins, you know, lipid lowering drugs. Didn't they begin testing them in individuals who were – had a

genetic marker, were familial hyper hypercholesterolaemia. And they found the effectiveness there and then they brought it out to the larger community, am I incorrect on that? If no one knows then that's what happened.

John Dwyer: The answer is I happen to know this, the answer is, you're predominantly right, there were a couple of other categories that they went and – if you read the original label and there were a couple of other categories, and Howard, you'll recall, there was this thing called off-label use, and it's perfectly appropriate in certain circumstances and you got a lot more scripts than the particular original claims might have warranted, okay. That's not gonna happen in the future. You're gonna have very narrow reimbursement and it's gonna be slaved by how you size your claim is my initial proposition.

And so the two things work in tandem and since you let me comment, and John, is anybody doing research like this because –.

John Dwyer: Yes.

John Dwyer: Yes. It has not been – it's been disappointing in terms of the information that we get. That doesn't mean that we won't be able to detect a positive signal but to date it just has not been as informative.

Dale Schenk: Yeah, the comment I always have about blood tests, having worked on that for a decade is, it's a little bit like looking in CFS for cholesterol changes. The last I checked, Alzheimer's disease is a brain disorder and to look in the blood just doesn't make sense. Now, that doesn't mean we shouldn't do it and shouldn't continue to do it, but wanting it and having it work just hasn't happened, and I doubt it will.

John Dwyer: This is now the John Dwyer show. There is a allele test right now in blood, where they're finding tumors out of the blood test. So I hope that some antigen can be found.

Dale Schenk: Well, with oncology, you can have a breach of the blood brain barrier, and so that's why that works. But, you know, in Alzheimer's disease it's not so clear there is. So, wanting it and having it be true is just – I mean I honestly spent a decade – myself and my colleagues and many of us here looking for it, and it's a different compartment.

Dan Perry: I know Eric Siemers had a comment.

Eric Siemers: Well, yeah, just a quick comment about how acceptable screening would be with lumbar punctures. I guess I'll just quote Dave Holtzman, because he always uses this line, which would you rather have the lumbar puncture or colonoscopy, so, you know, from my standpoint it's pretty clear.

But seriously, I think in terms of payment for the diagnostic tests, if you go to Europe and talk to the reimbursers there, and of course, by the end of next month we'll have a whole new healthcare system here and maybe it'll look the same, but there's less cost shifting there. And so you look at the people who model the health economics of these treatments, they just figure in the cost of the testing. It's really not that difficult to do. And then at the end of this, you're gonna give them X amount of data and they're actually gonna extrapolate beyond the data from a health economic standpoint and they'll just do the math. And you have to make the math work but it's doable.

Dan Perry: We have time for a few more questions, but I'd like to extend the invitation to Dr. Katz or any of your colleagues from FDA, to pose a question, instigate further discussion amongst any of those in the front of the room, if you have any. The FDAers? All right, let me go to over here.

Sue Peschin: I'm Sue Peschin with the Alzheimer's Foundation of America. I just wanted to follow up on sort of the side comment that you made, Dale, about mixed pathology, because it does seem like the neuropathology community has been coming out recently with more articles on how dementia tends to be much more of mixed variety than previously thought, and that, you know, it seems like there is sort of growing chorus that that's gonna eventually be recognized as the most common type of dementia.

And I just wanted to know, I guess I didn't hear very much about it today. I know Ron sort of mentioned starting to talk more about pathology, what the impact of mixed pathology was on all the work that's being done on biomarkers and on disease modifying therapies. If people could talk about that a little bit more.

Dale Schenk: Yeah, I mean we all could.

Dan Perry: Who'd like to take that?

Dave Knopman: Well, I think that it is a very critical area. I certainly agreed with the observations. A couple of comments. If you – we haven't talked any here about the differential biomarker effects with advancing age. In fact, there probably are some. On the neuropathologic level, the associations between cognition and burden of neuropathology actually attenuates once you get into the tenth decade and presumably, that's because there's other pathology and presumably a bulk of that other pathology is cerebrovascular.

That said, it still is important to focus on specific pathologies that you can do something about and so we're focusing here on Alzheimer's, but as many – as other people have said, that doesn't mean that simultaneously, approaches that are geared at cerebrovascular disease couldn't be undertaken as well.

However, it's my view and others may disagree and I can't prove it, if you look at the cardiovascular risk factor data, where it makes a difference is in midlife. In fact, cardiovascular risk factors sometimes flop in old age, that metabolic syndrome or obesity, which are risk factors in midlife for later life dementia become protective and actually it's true of hypertension as well. And it's a little complicated, but the focus is in midlife and there are things we can do about that now that aren't rocket science. But it's – unfortunately, those are huge societal weaknesses like obesity and dietary issues, smoking.

Dan Perry: Others on the panel? Any other questions from the audience? Marc, Dr. Walton.

Marc Walton: A comment or two that I'm wondering if perhaps the panel would have any thoughts about. In the conversations that we've been having here across the entire morning, there is a couple of times where seeming conflicting needs have been brought out of the issue of the outcome measure of the very individualized functional outcome measure versus a nonspecific global measure versus a multi-component global measure that is not functional.

And the issue of sensitivity of comprehensiveness, the issue of being convincing to the regulators for approval, for marketing, the issue of being convincing to the patients and physicians for using the product and for payers for paying for the product. And how the different outcomes measure satisfy – it's difficult to find outcome measures that are going to be able to achieve that.

Similarly, on the eligibility criteria, there is conflicts between identifying a group of people who have the disease and are going

to have development of enough symptomotology or progression of symptomotology fast enough to demonstrate the effect and how that may narrow the group of people who can ultimately get it, versus a broader population for whom the – presuming you show it works, that the broader population is then eligible to get it, either medically eligible or financially eligible.

And it strikes me that perhaps a useful thing to do is to think about not the issue of designing a clinical trial and what are we using as the outcome measure in the clinical trial or the eligibility criteria in the clinical trial, but rather to think about what's being done in the drug development as a whole. It's a drug development program, that when there are multiple objectives to a clinical trial, and these days they all have multiple objectives of who the clinical trial is going to convince for some particular decision that that examiner, whether it's the third-party payors, the physicians and the patients or the regulators, their decision to make, maybe they need different outcome measures.

And therefore, when designing a clinical trial that that be considered and there are multiple measures incorporated so that each of the needs can be satisfied but primarily by a different outcome measure. Similarly, when thinking about the eligibility criteria, and the needs of the population as a whole, think of it not as a clinical study that's going to establish everything but a clinical study that is going to establish an initial component of what we need and how that is designed to lead into other clinical trials that will provide the other elements of what we need, such as broadening the patient population.

For instance, the first clinical trial may establish that in people who have symptoms and have some biomarker that illustrates the rapidity of progression, and establishes efficacy and safety in that population, is then designed to easily lead into subsequent clinical trials that may be the low symptomatic or the nearly – demonstrably impaired but not disabled patients and thereby can bring them into the efficacy population. Or that the first clinical trial uses the biomarker to select but having done that and shown impact on the biomarker, that that establishes the ability to take people who have lower levels of the biomarker and less clinical symptomotology and combine between an effect on a biomarker and minimal effects on symptomotology. Establish that it is the broader population that can benefit.

So, basically just the idea of multiple outcome measures for different objectives may be necessary and may be useful to bear in

mind and the development program as a whole, planned to have multiple trials to satisfy the needs.

Dan Perry: Wow. I want to thank you for adding another layer of complexity. Who would like to respond to that on our panel?

John Morris: Jeff, that should be you.

Jeff Cummings: I think this is more of an industry question, but I mean I think you're exactly right and I think you've articulated what many drug development programs incorporate, which is trials that have different goals. It is hard to have enough outcome measures in on trial to answer all of the questions of the different audiences that drug development must address.

So, for example, phamacoeconomic outcomes in classical trials are often difficult to interpret because the population is so tightly defined for population entry. So I think what you've articulated is pretty much what is done, which is multiple trials to address multiple different audiences. Dale –?

Dale Schenk: Yeah, I'll say one or two words and then maybe Eric or Ron will want to say something too. But yeah, I tried to get at this with the question I asked earlier but I didn't do a very good job at it. I mean the first step, of course, is a given agent safe and effective and the part I think you're talking about is effective.

If we can show that a given agent is efficacious in a design phase three, then all of these other questions enter into it, and I think Rusty made the appropriate comment that for some of the disease – potentially disease modifying agents that arte being tested today, we don't quite know yet what that will mean on all these other measures that you're referring to.

And so it is gonna be incumbent upon not just the sponsors but I think the field in general to dig in and figure it out, because it will affect the way that we treat patients, it will affect society, it will affect all these other parameters. And we might find again, different classes of agents will benefit different downstream consequences of the disease and we'll have no way of knowing it if we don't do those type of studies you're alluding to.

Dan Perry: Dr. Black.

Ron Black: Point out that I think that what you outline is certainly is what we all aspire to, and I could just say some of my colleagues at Wyeth

are working on a drug development program for a drug that's for a basically opioid induced constipation and the outcome measure is three days later, and you can do all these things. You can actually do some really nice adaptive designs and you know, broaden the indication. One of the problems with what we're doing is that it seems pretty clear that the outcomes happen in 18 months, you know, and so we have a tendency to load our trials up with maybe too many objectives so that we start to lose some of the scientific focus.

But it's just hard when the outcomes, that you have a slowly progressing disease with a lot of variability and the outcomes are so far off in the distant future, you know, at least on the scale of drug development in other areas.

Dan Perry: Others on the panel?

Eric Siemers: I mean just briefly, I think – maybe I could summarize your question, it's how much do we want to be lumpers versus splitters is sort of what it boils down to, and you know, I think the answer is usually somewhere in between. We do have a lot of outcomes measures and it's possible that some people have some outcomes vary more than others, especially in terms of where they are in the disease course.

But that's, I guess part of the art of clinical trials, is to make them doable but at the same time capture a broad enough base of information that you've really described the effect of the treatment.

Dan Perry: Another question from the audience before we move to a summary? Anyone? Okay. Let me take just a moment to share a couple of observations and a little bit of looking back over the last four hours and comment on what I think we've accomplished.

First of all, let me again express a deep sense of gratitude for all that have made today's meeting possible, and I start by acknowledging our co-hosts, the Alzheimer's Association and the Leaders Engaged on Alzheimer's Disease organization. Between those two groups and the 50 not-for-profit organizations that are part of the ACT-AD coalition, you have given us the wherewithal to really bring the larger community to stand at the crossroads of clinical medicine and regulatory science, and I think that is a very special achievement. It couldn't have been done without all of those organizations joining us shoulder-to-shoulder to help make this possible.

I want to thank the sponsors who have allowed the noncommercial, not-for-profit organizations to have the wherewithal to carry out a meeting like this and to be able to back us in playing this very important role at this confluence of science and regulation.

Obviously, deeply indebted to the clinicians, the scientists, the clinical trialists who have – really the crème de la crème of the field, who have come back for today's meeting, have shared their thoughts and their insights with this audience. You have established an incredibly high level of expertise, which frankly, I'm in awe of. And so grateful to all of those.

And a special gratitude to the Food and Drug Administration and to its leadership. You've come here truly with open minds, I believe, and with open hearts to engage in this broader community. You've shared your thoughts. You haven't held back. You've been frank, direct. Have helped really forge an honest exchange of ideas. And you've shown a great deal of flexibility and I think that encourages us that this is not – will not just go down as another meeting, but perhaps a turning point, as we try to bring science to bear to find ways to move forward on these important issues.

And I also want to observe what so many have said, this is hard. This is extraordinarily hard. We're talking about a disease that is still in some senses ill-defined. If it's not more than one disease it is, as Dr. Rockwood and others have said, patients differ and they show progress in different ways. We've even seen that some in the placebo group do better than those in the experimental group and so forth. So it's devilishly difficult but that doesn't mean that we can stand back and let nature take its course and wait 10, 15, 20 years for the smoke to clear and to move ahead. So I appreciate very much those that have helped dig in on such a difficult issue.

And we opened it up with Dr. Rockwood putting down a rather, I thought, provocative marker about individualized disease, and I think you got a very positive response from many of the members of your panel as well as from the Agency, saying that there is attractive aspects to individualization and goal-attainment strategies that look at what benefits a particular patient in his or her circumstance, as well as their caregiver.

And the Dr. Cummings took the wheel with setting down a basic quantitative measure that we might agree would indicate a minimal clinical benefit that we could all rally around. And I think Dr. Katz found – or excuse me, Dr. Cummings found himself pushing on an open door, because Dr. Katz very readily said, we don't have

any number that you have to reach, and in fact, alluded to a package of indications, which could include clinical biomarkers, it could include global measures and include cognitive testing as well. And that there is no dogma about what numbers you have to hit and which one of those, they're going to stand back and look at it overall and take the larger picture. I think that was enormously gratifying for all of us that want to see the Agency willing to work with it, go with the emerging information.

And then David Knopman and his panel carried us even further along the road of discussion of biomarkers at how they might, at a very minimum, help us design better clinical trials, cleaner clinical trials, where we know we've got a population that really has the disease that we're trying to affect. And that gets us that much closer to the day when we will have therapies that have been tested in this way and that are ready to benefit our people and indeed, people around the world.

So, I began with gratitude and I was encouraged by your willingness to tackle a hard issue, and ultimately, end with a high degree of hope that this will be a turning point meeting, and in the weeks and the months ahead that we'll continue to feel some of the reverberations in the interactions between the clinical community, the clinical trialists, industry, patient groups and the Agency.

So again, I thank you for all of that. We've often said that this meeting was to be the deeper dive on the meeting that we began on clinical meaningfulness back in 2008. I think you've hit just the right level of the dialogue. This has not been one presentation after another. You've really dug down into the issues. I think there's been just the amount of new data presented, not overwhelming, just the right amount, it seems to me. Just the amount of earlier studies, even some that are just about to be published that have not.

I think there was just enough disagreement that you know that this was not cooked, and I just feel that we hit just the right level. And again, I'm enormously appreciative of all of you for participating and thank you very much and enjoy the rest of the day.
(Applause.)

[End of Audio]