00001

CENTERS FOR MEDICARE AND MEDICAID SERVICES
Medicare Evidence Development & Coverage Advisory
Committee

May 21, 2008

Centers for Medicare and Medicaid Services
7500 Security Boulevard
Baltimore, Maryland

00002

Panelists

Acting Chair
Saty Satya-Murti, M.D., M.Sc.

Panel Members
Marion Danis, M.D.
Daniel D. Foley, M.D.
Mark D. Grant, M.D., M.P.H.
Spencer H. Kubo, M.D.
Stephen L. Ondra, M.D.
Stephen Pauker, M.D.
Andrew Sloan, M.D., F.A.C.S.
Jonathan P. Weiner, Ph.D.

Patient Advocate
Leslie B. Fried, J.D.

CMS Liaison
Louis Jacques, M.D.

Consumer Representative
Randel Richner, B.S.N., M.P.H.

00003

Panelists (Continued)

Industry Representative
Jose Alvir, Dr.P.H.

Guest Panel Members
Naomi Lynn Hurwitz-Gerber, M.D.
Elliott J. Roth, M.D.

Executive Secretary
Maria Ellis

00004

TABLE OF CONTENTS

00005

CONTENTS (Continued)

00006

PANEL PROCEEDINGS

(The meeting was called to order at 8:15 a.m., Wednesday, May 21, 2008.)

MS. ELLIS: Good morning and welcome, committee chairperson, members and guests. I am Maria Ellis, an executive secretary for the Medicare Evidence Development and Coverage Advisory Committee, MedCAC. The committee is here today to discuss the evidence, hear presentations and public comments, and make recommendations concerning the design and methodological issues that challenge clinical research regarding innovative neurorehabilitation techniques. The meeting will discuss the various kinds of evidence that are useful to support requests for Medicare coverage in this field.

The following announcement addresses conflicts of interest issues associated with this meeting and is made part of the record. The conflict of interest statutes prohibit special government employees from participating in matters that could affect their or their employers' financial interests. Each member will be asked to disclose any financial conflicts of interest during their introduction. We ask in the interest of fairness that all persons making statements or presentations also disclose any

00007

current or previous financial involvement in a company that manufactures or provides devices or other tools for the research of innovative neurorehabilitation. This includes direct financial investment, consulting fees and significant institutional support. If you haven't already received a disclosure statement, they are available on the table outside of this room.

We ask that all presenters please adhere to their time limits. We have numerous presenters to hear from today and a very tight agenda and therefore cannot allow extra time. There is a timer at the podium that you should follow. The light will begin flashing when there are two minutes remaining and then turn red when your time is up. Please note that there is a chair for the next speaker and please proceed to that chair when it is your turn.

For the record, voting members present for today's meeting are: Marion Danis, M.D., Daniel Foley, M.D., Mark Grant, M.D., Spencer Kubo, M.D., Stephen Ondra, M.D., Stephen Pauker, M.D., Andrew Sloan, M.D., Jonathan Weiner, Ph.D., and Leslie Fried, J.D. a quorum is present and no one has been recused because of conflict of interests.

The entire panel including nonvoting

00008

members will participate in the voting. The voting scores will be available on our web site following the meeting. Two averages will be calculated, one for the voting members and one for the entire panel. I ask that all panel members please speak directly into the mikes. You may have to move the mikes since we have to share. If you require a taxi, there is a sign-up sheet at the desk outside of the auditorium. Please submit your request during the lunch break. And lastly, please remember to discard your trash in the trash cans located outside of this room.

And now I would like to turn the meeting over to Dr. Louis Jacques.

DR. JACQUES: Thank you, Maria, and thank you all for coming today. This clearly is a very broad and complicated topic for a number of reasons, including the fact that the underlying disease itself is quite complex. If we manage to completely solve everything to everybody's satisfaction today we will have exceeded my expectations tremendously. If we manage to make some progress and enlighten ourselves and each other on this, then I think that is certainly a reasonable goal.

To my immediate right chairing the

00009

committee today is Dr. Saty Murti.
DR. SATYA-MURTI: I'm Dr. Saty Satya-Murti, I am a neurologist and a consultant for sometimes CMS and sometimes industry, and I have no conflicts of interest for this particular meeting, and I'm not representing any neurology societies or academies.
I wanted to ask Maria, do I get to vote? No, okay. I have been on these panels before but I just wanted to be sure.
DR. DANIS: I'm Marion Danis, from the Department of Bioethics at the National Institutes of Health and run the ethics consultation service there. I have no conflicts of interest.
DR. FOLEY: I'm Dan Foley, I'm an emergency physician and I'm the medical director of Allina Health System. I have no conflicts of interest.
DR. GRANT: I'm Mark Grant, I'm an associate director of BlueCross BlueShield Association's technology evaluation center and I have no conflicts of interest.
DR. KUBO: My name is Spencer Kubo, I'm a cardiologist from Minneapolis-St. Paul. I'm also global medical director for Acorn Cardiovascular. I

00010

have no conflicts.
DR. ONDRA: I'm Steve Ondra, I'm a professor of neurological surgery at Northwestern University. I am a consultant to Medtronic and receive research grants, but nothing in this area.
DR. PAUKER: I'm Steve Pauker, I'm from Tufts, I have no conflicts, although sometimes I wish I did have some.
(Laughter.)
DR. WEINER: I'm Jonathan Weiner, a professor from here in Baltimore at Johns Hopkins University. I have no direct conflicts of interest, although within the Johns Hopkins University, I'm sure among my 15,000 colleagues there may be.
MS. RICHNER: I'm Randel Richner, Neocure, a private consultant on health economics and reimbursement. I have no conflict in this particular issue today.
DR. ALVIR: I'm Jose Alvir, I'm the industry representative. I work for Pfizer. Pfizer does not have any tools or devices for this particular issue, although we do have drugs for neurologic disorders.
MS. FRIED: I'm Leslie Fried, I'm employed at the American Bar Association Commission on Law and

00011

Aging, and I direct a joint project with the Alzheimer's Association on Medicare coverage issues, and I have no conflicts of interest.

DR. HURWITZ-GERBER: I'm Lynn Gerber, I'm the director of the Center for the Study of Chronic Illness and Disability at the George Mason University in Fairfax, Virginia. I'm on the board of governors of the Academy of Physical Medicine and Rehab but I'm not representing them today and I have no conflicts of interest.

DR. ROTH: Good morning. I'm Dr. Elliott Roth, I'm chairman of physical medicine and rehabilitation at Northwestern University School of Medicine and chief academic officer at the Rehabilitation Institute of Chicago and I have no conflict of interest.

MS. STILLER: Good morning. I want to thank you, chairman, panelists, invited guests and members of the public. On behalf of the Centers for Medicare and Medicaid Services, welcome to today's MedCAC meeting on methodological issues and evaluation of innovative approaches to stroke rehabilitation.

I would like to take this opportunity to introduce myself and the CMS analytic team

00012

responsible for today's meeting. My name is Jean Stiller, and my role is lead analyst for the project. Stuart Caplan is a senior analyst on the team. Dr. Susan Miller is the lead medical officer. And Maria Ellis, who most of you already know, is the MedCAC executive secretary. Dr. Louis Jacques is the director of the Division of Items and Devices and Dr. Steve Phurrough is the director of our Coverage and Analysis Group.

There are two main goals for today's MedCAC meeting. The first goal is to clarify the design and methodological issues that challenge research in the field of neurorehabilitation, specifically in the area of stroke. The second goal for today's meeting is to identify the desirable characteristics of research trials in this arena.

Dr. Susan Miller will kick off today's events with a presentation that focuses specifically on the characteristics associated with neurorehabilitation research, using stroke as the incident disease.

Next we will hear a presentation by Dr. Mark Oremus and Dr. Pasquilina Santaguida of the McMaster University Evidence-Based Practice Center. You will hear the details about the research they

00013

conducted in response to the technology assessment commissioned by the Agency for Healthcare Research and Quality. The technology assessment is one of the primary inputs used by the panelists to formulate recommendations on today's topics. Panel members were also provided with additional background materials determined relevant to the subject matter. Finally we will hear several presentations from invited speakers and interested parties.

Questions posed to the MedCAC panel consist of voting and discussion type questions. For those questions in which panelists are asked to express a degree of confidence, individual panel members will be asked to respond with a score from one to five; a score of five indicates that a panel member is very confident in response to the question posed, whereas a score of one indicates a complete lack of confidence for that particular response. Discussion type questions are not scored but allow for a free exchange of ideas in the area surrounding that particular topic.

I will now read aloud each of the 11 questions that the panel will later react to by either casting an individual score in the case of the voting type questions or discussing in detail for the

00014

case of the discussion questions. Out of the 11 questions posed, nine questions will be scored. Two questions, number 2 and 11, are for discussion purposes only.

There is the tendency to generalize stroke research to large heterogeneous populations. How confident are you that the strategies below represent meaningful comparators in observational studies? A, protocol-driven usual treatment versus protocol-driven usual treatment using the same parameters plus the specified intervention. B, patient him/herself before and after intervention. C, patient him/herself before and after treatment, then with treatment withdrawn and reinstated as appropriate. D, non-protocol-driven usual care versus intervention.

Panel Question Number 2: Large prospective randomized trials are uncommon in this field of medicine. Discuss how other study designs can or cannot adequately account for potential confounding factors such as: A, natural clinical course of recovery. B, selection bias due to skill level of therapist, comorbidities affecting both the stroke etiology and course of recovery, ancillary therapeutic resources, virtual home/community

00015

environments, severity of illness. C, differing assessment tools used across care settings, inpatient rehabilitation facilities, skilled nursing facilities, home health agencies, outpatient centers. D, premorbid and cultural characteristics. E, discharge settings and social support.

Panel Question Number 3. What is the minimum period of time that interventions be followed in order to identify a durable treatment effect? A, zero to six months; B, six to 12 months; C, 12 to 18 months; D, greater than 18 months.

Panel Question Number 4. How confident are you that each of the following outcome measures is a reliable, valid and responsive indicator of change in clinical trials that aim to improve an individual's functional capacity in the performance of ADLs, IADLs and locomotion/transfer abilities? A, Barthel Index; B, six-minute walk; C, functional independence measure; D, Fugl-Meyer Assessment.

Panel Question Number 5. How confident are you that each of the following outcome measures is a reliable, valid and responsive indicator in clinical trials of therapies to improve an individual's functional capacity in the performance of language and communication skills? A, Aphasia

00016

Quotient of the Western Aphasia Battery; B, Porch Index of Communicative Ability.

Panel Question Number 6. How confident are you that each of the following outcome measures is a reliable, valid and responsive indicator in clinical trials of therapies to improve an individual's functional capacity in the performance of swallowing? A, coughing/choking frequency during a meal; B, video fluoroscopy.

Panel Question Number 7. How confident are you that each of the following outcome measures is a reliable, valid and responsive indicator in clinical trials to assess patient, proxy, or caregiver perceptions of the patient's health and satisfaction with life and community reintegration? A, Barthel Index; B, Modified Ashworth Scale; C, EuroQol, quality of life for patient and caregiver.

Panel Question Number 8. How important are caregiver burden and their narratives as indices of successful rehabilitation?

Panel Question Number 9. How confident are you that these conclusions can be generalized to community practice settings outside the context of specialized treatment centers?

Panel Question Number 10. How confident

00017

are you that these conclusions can be generalized to the population of Medicare beneficiaries? Panel Question 11. What are the gaps in the current evidence on stroke rehabilitation therapies and Medicare beneficiaries? Finally, I would like to introduce my colleague, Dr. Susan Miller, who is the physician lead medical officer on this project. Susan is board certified in physical medicine and rehabilitation and has been in community and academic settings for over 20 years.

DR. MILLER: Thank you, Jean. Good morning to all and welcome to today's MedCAC. We hope that you find this meeting to be a productive one. As the field of neurorehabilitation is evolving, CMS is constantly being requested to consider new technological devices for coverage consideration. When faced with these questions, CMS considers, does the new technology, be it a device, a procedure or a therapy, produce a clinically significant benefit? Does it do more good than harm? And are these answers generalizable to our beneficiaries, who are of course those who are over 65 years of age, those who are disabled and those with end stage renal disease.

00018

In order to answer these questions, CMS takes an evidence-based approach to the decision-making process and looks for quality research which in general we believe should include at least a specific clinical question to examine, a study design that will answer that question well, both defined interventions and comparators, appropriate measures of outcome, confidence that the execution of the study promotes a truthful answer to the question, and of course, again, useful answers that are applicable to our population group. The field of neurorehabilitation technology is too large to cover in today's MedCAC and so we have chosen stroke and its associated technology as our incident disease. We thought it made sense to use stroke because in the United States alone, approximately 780,000 individuals experience a stroke each year. Three-quarters of these strokes occur in those who are 65 years or older. 600,000 of these strokes are primary attacks, 180,000 are recurrent. Stroke is not only a leading cause of death in this country, it is also an incredibly disabling disease. Among those who are 65 years or older and are six months out from their stroke, it has been estimated in at least one study that over 30

00019

percent require some sort of assistance in ambulation and 20 percent were dependent in some way in their activities of daily living.

A stroke occurs when the brain is deprived of oxygen and other nutrients. The most common type of stroke is known as an ischemic one. Ischemic strokes happen mainly when a blood vessel in the brain is clogged off by usually a combination of atherosclerosis and blood clots. In a hemorrhagic stroke a blood vessel bursts, causing damage by either leaking blood around or into the brain.

In either case, depending on where in the brain the harm is done, the effects of a stroke can include one or a combination of the following: Weakness or paralysis, usually on one side of the body; sensory abnormalities, again usually but not always on one side of the body; spasticity, meaning that the arm or leg can become very rigid and even move involuntarily; dysphasia or impaired swallowing; neglect, lack of awareness on usually the left side of the body and of the environment on that side.

Strokes can also cause communication difficulties, including trouble speaking or comprehending language. Strokes cause visual disturbances, loss of control of bowel and bladder, and they also cause cognitive

00020

impairment, meaning that there are difficulties with attention, planning, reasoning, problem solving, and very importantly, learning. A stroke can cause one to have balance difficulties, emotional and behavioral changes. Stroke is often associated with depression and impulsivity. And strokes can also cause spatial perception impairment, meaning that there can be difficulties in a person's ability to judge distance, perhaps they will also confuse right or left, or have trouble figuring out how to put on their clothes.

It is estimated that there are more than five million stroke survivors living in our country today, and though there are many technologies out there designed to improve the capability of individuals who have experienced a stroke, CMS must determine if there exists a clinically meaningful benefit to the various medical technologies presented. A good part of this decision is based on the review of the data that is found in the medical research literature. Therefore it is important to ask, what are the key methodologic issues that challenge research in the field of neurorehab in general and strokes specifically.

In order to help answer this question CMS

00021

commissioned a technology assessment to review and critique the current stroke literature. The TA, as it is otherwise called, concentrated not on the treatment methods found in the literature but instead on the methodology used to study those treatments, attempting to identify the more robust characteristics of quality research investigations. The TA looked at the stroke literature to analyze these datapoints that you see here, which we all agreed upon to be important in the evaluation of research in this field. As we have a number of persons in our audience today who do not do research, let me briefly define and discuss each of these points for you. The study design describes the approach that the research will take to the question that is asked. Now when a new technology is being tested it is compared against, for lack of a better term, the old way of doing things. Oftentimes the type of study design used for clinical research such as this is a randomized controlled trial, where at least two groups of patients are chosen for investigation, one group uses the old technology, the other uses the new, and at the end of the studies the groups are evaluated to see which did better. Theoretically if the group

00022

using the new technology did better than the one that didn't, then we should think that the improvement seen was because of the new technology, but this is not always the case, as we will discuss. Nonetheless, because the randomized controlled trial is so common in the literature, we have chosen to examine its characteristics today. We have also chosen to look at one of its cousins, the systematic review. Systematic reviews are a type of study design in which a literature search of relevant articles on the same specific topic are performed and then analyzed for predetermined key characteristics. If these specific characteristics are found, it allows the results of many different small studies, again, all on the same topic, to be combined, summarized and interpreted as a larger whole. Frequently systematic reviews search randomized trials for this purpose and so we are discussing both types of research design today. Before I go on, however, I do want to emphasize that there are numerous other study designs that may be applicable to medical research, some of which are noted here. That we are using randomized controlled studies and systematic reviews today as our discussion platform does not negate the potential

00023

of other appropriately chosen study designs to also provide answers to our questions. The use of randomized controlled trials may very well be unsuitable in certain indications, and as you just saw, today our panel will be discussing other research design strategies.

The larger point that I am making here is that study designs and all the other specific examples that are used to illustrate our data points today have their own unique advantages and disadvantages that make them applicable or not to various situations. CMS does not endorse design methods, outcome measures or the like as appropriate to use. Instead, our discussion today is to focus on those over-arching design and methodologic concerns that we all need to bear in mind in order to make the best coverage decisions possible for our beneficiaries.

Now having said that, let's return to our example, the randomized controlled trial. One of the reasons researchers use randomization is that if it is applied properly, it lessens bias. Bias is anything that can affect the results of a clinical investigation, anything besides the intervention that you are studying. In the randomization process the

00024

investigator is trying to make his or her two separate groupings of people as comparable as possible, so that at the end of the clinical trial, if one group is found to have a better outcome than the other, we should again be able to logically conclude with some confidence that one intervention is better than the other.

In many of the studies that we see at CMS the research groups are randomized, usually for age and gender. However, people who have had strokes are a heterogeneous or diverse population. Therefore in a randomized controlled trial, it may not be enough to just randomize only for age and gender. You may have to consider other factors like the etiology or the cause of the stroke, the type of injury caused by the stroke and the functional and emotional consequence of that injury. You may have to consider the health status of your subjects, their social and cultural environments, all depending on the characteristics which might affect the outcome of your study.

Consider for example that after a stroke, people commonly have trouble walking. The cause of this could be for one of many reasons. Some people can't walk because of either muscle weakness or

00025

paralysis. Others can't walk because of spasticity, others not because they can't move their legs but because they can't coordinate this movement with the rest of their body. Some can't walk because they can't understand the words that their therapists are using to try to communicate to them as they reeducate them in the skill. And some people may need more practice than they can receive in their therapy program, but they don't have family members or friends who can help them with this. And some people come to their stroke with multiple medical problems like heart and lung disease that limit their capability to walk again no matter how hard they try or how much they attempt to practice.

Any one of these or similar factors can influence the outcome of a study, yet be external to the actual effect of the intervention. Therefore, it is important to think of these factors at the beginning planning stages of research in order to be able to present unbiased results at its conclusion.

Now, suppose you have gone to all the trouble to make certain that your groups have been successfully randomized and therefore are comparable to each other at the beginning of your trial. You certainly would want to keep them that way throughout

00026

the entire study except of course for the intervention itself. And so the groups have to be treated equally during the trial.

We have noted that in the area of neurorehab research there are some different and perhaps unique types of factors that can cause difficulty in this area. Oftentimes technology needs to be taught to patients during therapy sessions. So some of the challenges that can factor into a neurorehab study result are represented by questions like, did the two groups receive the same type of therapy save for the intervention itself? Did they receive pertinent therapies in the same order? Did the study groups receive the same frequency, intensity and duration of therapy? Did the patients obtain their treatments and final evaluations from professionals with equal levels of skill and experience? Did they all perform their therapies in the same type of setting, and if performed in different institutions of the same type, did all the institutions in the trial provide their patients with the same resources? Granted, these variables are very difficult to tackle in a clinical trial, but do need to be addressed in some fashion to provide confidence that every attempt has been made to

00027

separate the impact of outside influences or unequal treatment from the actual results of the investigation.

Another factor that can undermine the quality of research by causing an unintentional difference between the two groups is that the individuals participating in the study may be at different points in time in terms of their recovery from their strokes. Physicians are aware that some patients can show the ability to recover either fully or in part spontaneously, meaning without medical help of any kind after a stroke, usually over the first six to 12 months.

So suppose for example more people who were recovering naturally were in the intervention group, as opposed to the group that used the old technology, and suppose at the end of the study the intervention group did better than the group using the old technology. CMS would ask, was the better outcome because of the new technology used, or would it have happened anyway because the people in the group were improving on their own. So to make certain that, again, the intervention is the only effect influencing the results of the study, investigators must consider an appropriate strategy

00028

in their study design to account for timing of recovery.

There can be many other sources of bias in a research study. One of the most devious is opinion. Most of us if included in a research study would come to the table with some preconceived notions. For example, it might be my opinion that because the technology is new or expensive or has been heavily marketed, that it just has to be better than the old way of doing things. If I then know that I am receiving the experimental intervention as opposed to the same old same old, I might just feel better because I expect to, that's just human nature. But it is, however, not a fair evaluation of the technologies at hand.

Therefore, consideration needs to be included in the study to reduce this sort of biased effect. Where possible, this can be accomplished through blinding. Blinding or masking, as some people call it, means that the people who have a stake in the study like the investigator and the patient do not know who is getting the new treatment and who is not until the study is completed. Again, this can help prevent an external influence like opinion from interfering with the actual facts of how

00029

the study results are interpreted. Admittedly, though, blinding is sometimes difficult to accomplish in the study of new technology, but the concept cannot just be rejected out of hand immediately. Again, patient selection criteria are also important to consider when planning a study. It doesn't make sense to include in your study groups people who are not typical of the real world population that will be expected to use the new technology if it is approved. Therefore, CMS needs to know the relevancy of the procedure presented to its beneficiaries, those who are 65 or older, those who are disabled, and those who are experiencing end-stage renal disease. Inclusion of these folks into any study submitted to CMS should be considered so that we can see the generalizable nature of your work to the Medicare population.

Another concern for CMS is when progress should be measured. CMS is mindful of the durability of any effect that a new technology might have. A durable study result is one that is relatively long lasting, it gives an idea to us of the usefulness of the technology. Many times study findings are measured at the beginning of the period of treatment, perhaps during that period, and finally at the end of

00030

the treatment period, and then never again. In this case we would have no information to evaluate these findings, to see if they are really useful to our patients not just while they're undergoing treatment, but also in their normal surroundings as they go about living their lives again. Durability of an outcome is a helpful means for us by which to evaluate clinical benefit.

The choice of a comparator is also very important in any study. What is a comparator? Well, consider that if a new technology is supposed to be better, then it's supposed to be better than what, and it is the what that is the comparator. The definition of an ideal comparator is the best available treatment in the field, given together with the best overall care of the patient. Because when undertaking a clinical trial of any study design, the investigators do so because they are reasonably uncertain if technology A or technology B provides the better service to the patient. After all, no patient should be subjected to a research group where it is already certain that one treatment is inferior to the other. So ideally the comparator should be no worse than the most effective treatments already on the market. To evaluate a new technology against a

00031

comparator already known to be substandard in its treatment abilities would make a study at its best greatly flawed, and at its worst, unethical. And as we are all trying to decide if new and improved is better than old, just how do we do that? What is it that is measured to demonstrate the success of a new technology? There are at least hundreds of outcome measures that have been used in the field of neurorehabilitation to gauge patient improvement or the lack therefor, but different outcome measures provide different types of information. How do you choose? Particularly in the realm of technology research, CMS is looking for outcome measures that describe a clinically relevant result. Clinically relevant is a difficult term to get your arms around. I just want to bring to your attention an example of a framework that can help you consider this concept as you go about choosing your outcome measures. The World Health Organization's International Classification of Functioning Disability and Health, or the ICF as it's known, considers three levels of functioning and the interconnected environmental background that are important in describing the health and health-related status of any given person. The

00032

three levels of functioning basically describe the bodily structure of the individual, the whole person, and then the whole person within his or her social context.

The first level of functioning called body functions and structures relates to just that, the physiologic processes and anatomic parts of our bodies and the changes that occur after our bodies are affected by injury or disease. Some examples of the components of this area are our mental processes, our muscle power, our muscle strength, our visual function, our joint mobility, range of motion, and the status of our muscle tone.

The second level of ICF functioning considers the person as a unit or as a whole and is pretty much defined by the activities or tasks we perform, whereas the ICF's third level of description, functioning of an individual within the context of society, is portrayed by the way we interact or participate in society, the roles we play among our family, our friends and our employers, for example.

These are a few examples of the activities and relationships that we all engage in that further define these two realms, the activities realm and the

00033

participation realm of the ICF. As you can see, they denote our basic activities of daily living like dressing and walking and toileting, as well as our communication skills and the relationships we have within our families and our communities. The ICF looks at all these levels of functioning, not only as a continuum but also as a set of forces that have effects upon each other. Different research questions might be considered as ways of moving between these levels of functional definitions. Which outcome measures are chosen for study, then, depends on the question being asked. Particularly in the realm of neurorehab technology research, CMS is mostly looking for outcomes and outcome measures that describe a result that is helpful to the patient by improving their ability to perform some sort of task that represents an activity of daily living. Usually, then, that outcome measure must be descriptive of a function that is related to personal care or independent living. So for example, it might be nice if you looked for, in a new device, you look for something that increases leg strength or joint mobility in a stroke patient, strength and range of motion being measures of bodily structure and function. But what

00034

might be really interesting is if this documented gain in strength or joint mobility was sufficient to allow the individual to walk further or to get on and off a commode, and do either or both of these two tasks with more independence and increased safety than before. So in your study it would be reasonable to choose outcomes that would measure the accomplishment of both bodily function and activities in the ICF framework.

My point here is that as you think about the outcome measures you will use in your data collection, it makes sense to bear in mind the interrelationships between all the health-related consequences of a disease process, and consider taking into account several types of measurements that have relevance to your study question and to the functional concerns of your patient population. Which outcome measures are used depends of course on the goals of the study, and would certainly be expected to be different for different research questions. But the outcome measures chosen do need to have scientific credibility, and that credibility should be authenticated in most circumstances for use in the population to which the measures will be applied. And certainly in larger populations of

00035

impaired individuals such as those who have experienced a stroke, these outcome measures should be authenticated specifically. Within that realm CMS will pay attention to these characteristics of the outcome measures at least. Their validity, the ability of the outcome measure to measure what it says it's supposed to measure. The reliability, the degree to which the measurement provides consistent and reproducible results when it's used in equivalent conditions. And the responsiveness of the measure, the degree to which it can detect change. We also will want to know that that change is not trivial but is important, significant and worthwhile to the healthcare status of the patient. These characteristics of outcome measures are known as their psychometric properties and they are absolutely essential to the demonstration of a useful purpose of a new technology.

I hope that I have now brought everyone up to speed as to why we are here today, as well as given enough background material to make today's presentations informative. Let me close by saying that an event like this does not happen through the efforts of only a few people, and I would like to

00036

thank my CAG team who have supported me through this endeavor. I would also like to thank our contributing team members from the Agency for Healthcare Research and Quality. I would also like to thank those who diligently worked on the technology assessment, and finally, last but not least, we thank our MedCAC panel members who today will recommend to all of us how to best improve and interpret the information that Medicare utilizes to enhance the healthcare outcomes of our beneficiaries.

At this point I want to say thank you to all of you, and I would like to introduce to you Dr. Mark Oremus and Dr. Lina Santaguida of the McMaster University Evidence-Based Practice Center. They will be presenting our technology assessment.

DR. OREMUS: Good morning everyone, it's a pleasure for both Dr. Santaguida and myself to be here today to speak to you about our technology assessment, and we'll get right into it here with a brief background. Dr. Miller had already mentioned some of these things so I will go through the next couple of slides very quickly.

In the United States there are approximately 600,000 incident cases of stroke annually, and strokes are the third leading cause of

00037

death in the U.S. They are also the second most common cause of disability. Two-thirds of the persons who suffer a stroke will actually survive the stroke episode but half of the survivors go on to have permanent disabilities.

The clinical consequences of stroke are variable and they are influenced by the location of the stroke in the brain and by the extent of cell damage, and the complications from stroke span a wide range of domains. For the purposes of our technology assessment we were interested in six of these domains which I have listed there, ambulation, quality of life, activities of daily living, cognition, communication, and dysphagia.

The efficacy of stroke rehabilitation interventions should be evaluated using evidence-based practice, and that is the use of the best available evidence to make decisions about patient care, and this maps into the methodologic flavor of this meeting, because obviously a strong research methodology is going to enhance the quality of the evidence.

So the purpose of our technology assessment was to evaluate the methodological quality of published studies in stroke rehabilitation and we

00038

used a series of eight evaluation criteria. Just to resolve any confusion, Dr. Miller had mentioned ten criteria, but on this slide some of those ten criteria are simply combined, so that's why we have eight criteria here, but it's the same as was outlined in the previous presentation. So our evaluation criteria included things such as study design, patient selection, randomization and blinding, and others which I won't go through right now but we'll talk about them as we progress through the presentation.

We decided to examine the published literature to assess the methodological quality of studies in stroke using two methods. The first method is called purposive sampling and the second method was a review of reviews.

For the purposive sampling approach we searched three medical databases over the last five to eight years to obtain up to 20 of the most recently published articles in our six domains of interest. I say up to 20 because in some instances we could not find 20 articles, and when we found more than 20 articles we decided to cap the number at 20, again, because our interest was in methodology and we felt that more recently published articles would be

00039

more likely to have more of the cutting edge, most up-to-date methodology used in this particular area. So how do we go about getting the articles? Well, we did our database searches and retrieved a certain number of citations. So we decided to screen these citations using several inclusion and exclusion criteria. We went through two levels of screening, the first being a title and abstract level. Articles that passed that level of screening went on to what we called full text screening. And articles that passed the full text screening level were abstracted, and we abstracted data on those eight evaluation criteria.

For our results of the purposive sampling we summarized the abstractive data into tables and charts and we also selected two studies from each of the six domains of interest and we provided in-depth summaries and descriptions of those two studies.

The second methodological approach we used in our technology assessment was called a review of reviews, and in order to conduct the review of reviews, we basically assessed the methodology of systematic reviews and the individual studies that were appraised within these reviews. To obtain systematic reviews we searched three medical

00040

databases over the last eight years to obtain relevant studies and the methodological quality of our systematic reviews and the studies in those reviews were summarized in tabular form.

So our results: For the purposive sampling approach, our initial literature search retrieved 1,674 citations, and 127 of those citations passed our first level of screening. Of those 127, 12 were duplicates, four were outside the range of the 20 most recently published studies in a domain, and one of the citations was just not retrievable, so 110 advanced to full text screening. And at the full text screening stage we excluded a further 11, so we abstracted 99 articles.

So now I'm going to go through a broad summary of what we found in terms of the methodology of those 99 abstracted studies. So in this slide here you can see that most of the 99 studies were randomized controlled trials. In terms of patient selection, all but two of the 99 studies reported both the inclusion and exclusion criteria used to select their sample, as well as rudimentary sample characteristics such as age and sex. In the quality of life and dysphagia domains, there was one study each that only reported one of those two categories,

00041

but at least all of the studies reported one. In terms of randomization, almost of the all of the randomized controlled trials indicated in the methods section that their studies were actually randomized, and there was 100 percent reporting in all of the six domain areas except for quality of life and dysphagia. In terms of blinding in the randomized controlled trials, approximately 75 percent of the RCTs reported that there was some blinding. However, in the cognition domain reporting was poorest with less than 50 percent of the authors of these studies reporting that there was blinding. Now looking at blinding in terms of both the randomized controlled trials and the observational studies which can include case control or cohort studies, approximately 75 percent of all of the studies described the type of blinding, and what I mean by type of blinding is they indicated who was blinded, was it just the outcome assessor or the outcome assessor and the patient. So now this slide will actually reveal to you what type of blinding was used in the study. So approximately 60 percent of the studies contained blinded outcome assessors, that is, the person who

00042

was assessing how the patient was performing was blinded. Patients, healthcare providers and data collectors were reported as blinded in less than 50 percent of the studies.

Approximately 80 percent of the studies identified the professional background of the person who was charged with actually delivering the rehabilitation therapy. In terms of the timing of the intervention, that is at what time post stroke is the intervention actually first delivered, the timing varied widely. Generally it fell within a range of zero to three months post stroke, but many studies did not report the timing of the intervention post stroke.

In ambulation, 100 percent of the studies actually did go and report the timing. For frequency and duration of intervention, a majority of the studies reported both the frequency and the duration of the rehabilitation therapy. However, half of the studies in the dysphagia domain reported neither. In terms of length of patient follow-up, again, they varied widely from study to study and from domain to domain. Typical follow-up was between one to 12 months. However in the quality of life domain, follow-ups tended to last longer than 12 months.

00043

Eight of the 99 abstracted studies contained no reported length of follow-up whatsoever. Regarding prior and concomitant treatment, slightly more than half of the studies did not report prior or concomitant treatment that might be relevant when you're assessing the efficacy of stroke rehabilitation therapy. This was the poorest reported key characteristic of the 99 abstracted studies.

In terms of the standard treatment comparator, virtually every study contained some report of details of the standard treatment and that's why, because reporting was so good in this area, that we did not give a graph.

For psychometric properties, we can see that in the ambulation domain there were 45 different instruments used in the studies of ambulation. Approximately 20 of these studies contained reports of whether the instruments had their psychometric properties in stroke. Now I have to clarify something here. When we assessed whether these instruments had psychometric properties in stroke, we looked at whether the authors of the abstracted studies or the authors of the systematic reviews reported whether the instruments they used actually

00044

had their psychometric properties in stroke. So approximately 20 in ambulation had psychometric properties as reported by the authors of the ambulation studies.

For quality of life, again, 45 different instruments were used to assess quality of life, and approximately 25 had psychometric properties in stroke. Very, very few of the authors reported details of whether the instrument they used had an established minimum clinically important difference in stroke.

In terms of activities of daily living, there were 25 different instruments used and approximately 15 were reported to have psychometric properties in stroke.

For cognition we're back up to that number of 45 instruments, but only three instruments were said by study authors to have had psychometric properties in stroke. I should also point out that more instruments may have had psychometric properties established in stroke, but we are relying on whether the authors of the studies indicated so. And none of the authors of the activities of daily living or cognition domain studies reported whether any of the instruments they used had established minimum

00045

clinically important differences in stroke. For the communication and dysphagia domains, there was absolutely no information presented by study authors on whether any of the instruments they used had psychometric properties in stroke. For ICF domains we looked at whether the authors of the studies reported what domains of ICF the instruments they used mapped onto, and when the authors of these studies made these reports, we found that the instruments they used tended to map onto one of three ICF components, function, activity or participation. So looking at the four domains where psychometric properties were reported for stroke, in the ambulation, quality of life, activities of daily living and cognition categories, we can see that a majority or a plurality of the instruments used were not mapped onto an ICF domain by the authors of the studies. For three of these four domains, activity was the most popular category of ICF when the instrument was mapped onto such a category, the one exception being in cognition where we only had three instruments where psychometric properties were identified, and so two of them mapped onto the ICF

00046

category of function.
I also talked at the outset of the presentation that we did an in-depth summary and review of two studies per domain. We found that when we reviewed these in-depth studies, 12 in total, we found that the methodologies of these studies generally coincided with the aggregative results that I just presented. So for the purposes of this presentation, I'm not going to present the results of the summaries of the specific studies.
So the next section of the presentation is going to discuss the results of the review of reviews, and my colleague Dr. Santaguida is going to take over and she's going to present the review of review results. She's also going to present our discussion and our conclusions.
DR. SANTAGUIDA: Good morning. So what we see on this screen here is a flow diagram which is typically presented in systematic reviews, and we see that we started off with 949 titles and abstracts which we initially screened to see if they were on topic. And from that, 204 citations were received and evaluated at full text, and you can see all the reasons why we excluded citations or publications at the full text level.

00047

Our final number was 36 systematic reviews of which 16 were Cochrane based and 20 were not. Within the systematic reviews, we again looked very carefully at how each of the reviews evaluated or critically appraised the studies eligible for each of the reviews. We looked very closely at what the criteria were that they evaluated and if they used a standardized checklist, and then where possible within the systematic reviews, if they provided individual study data with respect to quality assessment. So for example, if they evaluated ten randomized controlled trials on ten quality internal validity criteria, then we looked to see what proportion of those studies had changed the criteria that they evaluated.

What you see here is, on the X axis we see the different criteria, the quality criteria that we looked at for each of the reviews, and then on the Y axis we see the proportion of criteria that were either not evaluated within the systematic reviews or not achieved. So for example, with the dark colored bar graphs which are from the Cochrane reviews, we see that one category is empty, and that is for allocation concealment within the Cochrane reviews. That indicates that that particular quality criteria

00048

was evaluated in all systematic reviews from the Cochrane database, but we can also see that some criteria were not achieved or not evaluated in some of these reviews. For example, co-intervention and contamination was not well evaluated in either Cochrane reviews or non-Cochrane reviews. Similarly, adverse events. I realize it's a little bit of a busy slide but you can sort of see the idea.

So in looking at the systematic reviews we noticed that 38 percent of the Cochrane reviews did use a standardized checklist, of which the majority of these checklists had psychometric properties in the literature, and 78 percent of the non-Cochrane reviews used standardized checklists. One of the most frequently used standardized checklists was the PEDro scale which, and they used either the nine or 11-item version, and the PEDro scale is specific to randomized controlled trials, and you can see that the domains, the quality domains that are part of this particular scale, so for example there are three items relating to blinding, and there are two items relating to outcomes, and so on. The manner in which some of the studies that they evaluated within these systematic reviews achieved this criteria, as you can imagine, varied widely between the studies because

00049

there was a variety of stroke rehabilitation therapies that were evaluated within these systematic reviews. We also searched for reviews on outcomes specifically used within stroke and we found five such reviews that focused on outcomes in stroke. There was a range of studies that were included within these specific systematic reviews that varied from 32 to 357 included studies. The year of inclusion for these systematic reviews on outcomes also varied from 1966 to 2005. One of these reviews focused on acute stroke and the use of drugs. Three of these reviews looked at health-related quality of life outcomes, and one focused on all outcomes used to evaluate walking.

So, there were several methodological points to consider when evaluating studies in stroke rehabilitation. We believe that RCTs or observational studies, that is to say a study design that has a comparative group, are ideal for evaluating stroke rehabilitation. We believe it's important that the sample characteristics of subjects should be presented very clearly and hopefully in a table format and stratified by treatment group. We believe that the inclusion and exclusion criteria

00050

should be exquisitely stated within the study methods, and that the details of randomization and blinding, those details related to the potential for bias should also be clearly stated, reported in studies.

From the purposive sampling, again, here are some points that we think you should consider, that the authors should report in greater detail the type of professional delivering the therapy, the timing and frequency and duration of the intervention. They should report the length of the follow-up, prior and concomitant treatments. The comparator treatment, again, should be very clearly described. And of course, the outcome instruments selected to evaluate the rehabilitation therapy within the study should use measures that have psychometric properties established within the stroke population.

With regard to the review of reviews, there were a variety of stroke rehabilitation interventions that were evaluated within the systematic reviews, and the majority of these reviews did not restrict the inclusion of studies by the type of outcome. Usually the restriction was by the type of therapy that was being evaluated. Most of the

00051

reviews that included randomized controlled trials scored very high with regards to quality criteria for undertaking a good systematic review. The majority of reviews in terms of quality criteria that they evaluated within their eligible studies looked at randomization, blinding, withdrawals, dropouts. Fewer reviews evaluated very carefully or presented information very carefully about the comparability of the groups within the eligible studies, adverse events, co-intervention and contamination. And many of the reviews indicated that blinding of the patient and the provider was not possible for stroke rehabilitation.

With regard to the review of outcomes, a variety of outcomes have been used to evaluate the same attribute of interest, and from these review of outcomes specific to stroke it would seem that there is no single outcome that can likely capture all relevant dimensions of an attribute of interest, and that is to say that these attributes are conceptualized in quite complex ways. Also, there was a suggestion very much that if you're interested in evaluating a particular attribute, for example walking, then you should evaluate all components of walking, all components of that particular outcome of

00052

interest, walking outside, walking inside, walking on uneven ground and so on and so forth. It shouldn't just be walking in one dimension like walking indoors, in a rehabilitation study.

There was a very strong recommendation in the review of outcomes to select outcomes that had established psychometric properties, which we've discussed already, and of course to consider very a priori what would be a clinically meaningful change as opposed to a statistically different change.

Also, their recommendation was to consider what we called floor and ceiling effects, which are simply in part related to the attributes of the outcome that you select to measure what you're interested in evaluating, and also some very practical administration issues when choosing the outcomes.

The timing of the outcome measurement, again, should be justified, and some consideration of the time points in which you measure attributes within the patients, you should take into consideration the natural history of stroke recovery.

There are I'm sure several design challenges faced by researchers undertaking the evaluation of stroke rehabilitation therapies, and one of these of course is selecting and justifying

00053

the comparator treatment. You know, of course as Dr. Miller alluded, the comparator treatment should be one that is the best currently available relative to the treatment of interest, because if you choose something that you know a priori is not effective, that's clearly not a fair comparison. So selecting the appropriate comparator and justifying that is very important, and providing a theoretical rationale as to why the treatment and when the treatment may be having an impact in the recovery of stroke would also be very important in providing that justification for the comparator.

Very often in stroke rehabilitation there is a situation that we call multimodal type treatments versus unimodal. Clinically most rehabilitation therapies are what we would classify as multimodal, they have lots of small components put together that make sense clinically, and so sometimes this can present a challenge in terms of describing these therapies but also in evaluating them. There's some implications with the complexity of the therapy that should be better described and better justified.

Also, we recognize that a lot of times in stroke rehabilitation, although from a methodologic perspective you seek to standardize the therapy,

00054

there are situations where it makes sense clinically to adapt the treatment somewhat to a particular individual, and so when that has to occur in the stroke rehabilitation therapy, that that be better documented and better reported.

The other challenge I think that we determined from looking at the literature was, we did not notice that there was a consensus on how the timing post stroke was defined between studies. We think that there needs to be a better definition of what people mean exactly when they say acute, you know, subacute and chronic, because clearly we did not find a consensus with regard to this in the studies or the reviews that we evaluated.

We believe that there has to be a much better description of the care provider characteristics and possibly even the patient's provider interactions because that might be something very important depending on the type of stroke rehabilitation therapy that is being applied, with some consideration to reporting about the adherence to the therapy. There needs to be a better description of the system within which the care is provided. These are all aspects like Dr. Miller alluded to in the first presentation. And also the

00055

physical environment in which the physical rehabilitation is taking place. And if possible, a priori identification of subgroups who may respond differentially to the therapy.

We recognize that as researchers who look at numerous publications that we are always limited when we critically appraise a study in judging whether the researcher actually did something in their study design or in undertaking their study that limited the potential for bias, or that they simply didn't report it. And so for this very reason, within the research community and within the journal editorial community they have standards in which they ask people who are reporting their research to have a minimum amount of information.

And these, the examples that I have here is the CONSORT statement which is applicable to randomized controlled trials, and the STROBE statement which is applicable to observational studies. And what these statements indicate is the minimum amount of information that you need to indicate in your publication that would allow others reading your publication to know what you did or did not do to minimize bias. So again, if people adhere to the CONSORT or the STROBE, it makes it easier for

00056

us to determine if, you know, they eliminated bias, but it doesn't necessarily mean that they selected the right strategy to indeed do that in their research design.

This technology assessment had several limitations that we want to point out. One is that we only looked at publications in the English language. For the purpose of sampling, we selected comparative study designs. We selected a priori a subset of internal validity criteria to evaluate all the studies irrespective of the rehabilitation therapy that they used, and so we assumed that the therapy did not have an impact because we focused on the design criteria. Also as Mark mentioned, we assessed the psychometric properties of the outcomes in the pool of publications that we looked at based on the references that they provided within the publications themselves.

So what did we learn at the end of this technology assessment? Well, we found that many researchers did employ the randomized controlled design to evaluate stroke therapies and that many of these trials did have very positive, scored very well with respect to some of this quality criteria.

However, there were a few problems that we think

00057

researchers need to pay attention to, and inadequate reporting of the criteria that I have listed here, the randomization, the comparator treatment, the adverse events, inter-group comparability, we think that that can certainly be better reported. And again, we think that the justification for the selection of the comparator should be better described.

Based on the therapies that we observed in this technology assessment, blinding by these authors was consistently shown to be difficult to achieve, blinding of the provider and blinding of the patient. We also noted that ample size was sometimes an issue in some of these studies. Clearly if a study sample size is too small, it's very difficult to have what we call, you know, power as to the attribute which allows you to detect a change. And then many of the publications that we looked at also didn't really provide information about the minimally clinically important difference which is an attribute of the outcome measure, and helping us to understand the nature of the improvement that they measured. Also, there was problems with contamination and co-intervention.

Many of the outcomes reported in the

00058

publications that we evaluated did not have established psychometric properties within the stroke population and we think that that's very important. Also, many authors did not necessarily report the psychometric attributes of the outcomes that they selected for their study. We think that, you know, when selecting outcomes, we recognize that there's issues of practical administration that the authors need to consider, the validity of the self-report instruments, and also the rationale for the timing of when the measurements occurred.

So I think what we observed in the publications that we evaluated for this technology assessment is that we did find some good quality research for stroke rehabilitation therapies, but we still think there's room for improvement in some of the criteria that we've identified. Thank you.

DR. SATYA-MURTI: We have about five minutes for questions, it should be confined to the actual topic and not commentaries. I had a very brief question. We heard you say that randomization was not as good as we would have liked to see it and how difficult it is to randomize and blind, and yet two-thirds of your Cochrane reviews have good randomization. How did they overcome the

00059

heterogeneity of the stroke population in these randomization studies?

DR. OREMUS: We found at least in the purposive sampling section, which was my section of the report, that the randomization was just general randomization. And we almost had to come to that conclusion because although many of the published articles reported that the studies were randomized, that's all they reported, that there was randomization. They did not go in, the authors of these studies did not go into depth and indicate how they randomized, whether they took any issues into consideration when they randomized, was there some sort of a stratified randomization. They didn't go that far into the randomization, so it's very difficult for us to be able to assess how the heterogeneity of stroke populations was assessed via the randomization itself.

DR. SATYA-MURTI: So we have to -- almost to the end of your presentation, you said two-thirds of the studies had randomization, so I just wondered how sure can you be of that considering, again, just basic age, sex, base diagnosis, carotid occlusion, that's really not, you know, for stroke heterogeneity, is it?

00060

DR. OREMUS: Unfortunately one of the issue when we were evaluating methodologic quality of studies is since none of us were there when the actual study was conducted, we have to rely on what the authors report. So when we say that two-thirds were randomized, we're really going by what the authors said they did, and as I just indicated a few minutes ago, they really said we randomized but they didn't go into depth. So it's very hard to take what they reported in their methods section and make judgments about how they addressed heterogeneity via the randomization. It's a very difficult leap to make that assessment.

DR. SATYA-MURTI: Thank you.

DR. GRANT: You noted quite a few studies that by their nature lacked blinding. My question for you is, how much of a threat to the validity of the studies is your sense from reviewing those papers does that pose, how much potential bias, and what kind of efforts were made to account for that if in fact that was the case?

DR. OREMUS: In the 99 abstracted studies we found that many of the authors demonstrated a cognizance of the issues surrounding blinding, so that's certainly something positive to reflect upon.

00061

In general, attempts were made to blind the outcome assessor because that is perhaps the one area where an individual researcher could most easily address some of the biases that arise out of a lack of blinding, by trying to hide the assessor from what treatment the individual is receiving. At the same time they also indicated that in many stroke studies it was quite difficult to blind the patient or the person who was delivering the therapy. That's just a limitation, and of course there could be biases arising out of that. For example, knowing which treatment you are giving to someone and knowing the hypothesis of the study could influence how an individual may regard the treatment. But it's very difficult to, based on what was reported in the studies, for us to assess whether those biases actually had an impact on the results. That's, again, another leap that we can't make based on what we assessed. It's very difficult for us to assess how those biases may have actually impacted the studies. But there was, there were attempts to try to mitigate biases based on what we saw.

DR. PAUKER: I have two questions, both for Mark. How do you insure for the original sample, do you know if you see it in a few studies, how can

00062

you determine if it was not stated in the studies, and it should be in there if there were two people making the judgments or three people, how do we be sure that you have a valid study reported in there? That's question one. Question two, you reported a whole bunch of individual criteria, what is in every sentence, but how often do the studies, be it two, three, four, five, do they individually establish the criteria? If you could comment on both of those questions.

DR. OREMUS: Okay. For the first issue, that was related to study selection?

DR. PAUKER: Yeah.

DR. OREMUS: This report was a bit of an interesting report because unlike many systematic reviews that we would normally conduct, we weren't looking at efficacy, the mandate was to look at methodology. So our primary concern was selecting recently published studies in stroke rehabilitation so that we could examine them from a methodological perspective. So our funnel of selection was basically to assess whether or not the studies we captured in our broad literature research actually dealt with stroke rehabilitation, and we basically

00063

allowed any study into our assessment as long as it did deal with stroke rehabilitation therapy, again from a methodological perspective. We wanted to cast a broad net, and I was just going to say, as Lina had mentioned, that one restriction which was very important was that it had to be a comparative study. So if it dealt with stroke rehabilitation, if it was comparative in more instances than not it could be included. And we also looked at the most recently published studies because we wanted to get a sense of what was quote-unquote au courant with respect to the methodology.

DR. PAUKER: How many people made that judgment, one person, two? Was there any validation of when to exclude a study or was that one guy or two people, or both of you? How was that done?

DR. OREMUS: We had several screeners, approximately, I would say, how many would you say, Lina, at least six or seven screeners?

DR. SANTAGUIDA: Four.

DR. OREMUS: We had four screeners. Due to the volume of studies we had one person evaluate different chunks of studies, so there was no as you would say multiple validation of the same study, so it was basically studies were slotted and one person

00064

would evaluate.
DR. SATYA-MURTI: One last question if we have time. Dr. Ondra.
DR. ONDRA: This goes to the question about the way the studies were randomized and the way they balanced the populations appropriate to the question being asked and that is, how were they powered appropriately to the questions being asked?
DR. OREMUS: In many instances there were no sample size calculations provided either a priori or after the fact, so it was impossible to determine whether the studies were adequately powered.
DR. SATYA-MURTI: We should move on. There is, during the afternoon session at about 12:35, there will be opportunities for further questions. Sorry about that.
DR. MILLER: It is now my pleasure to introduce Pamela Duncan as our next speaker. Dr. Duncan is a professor in the division of physical therapy within the department of community and family medicine at the Duke University Medical Center. She is also a senior fellow in the Duke Center for Clinical Health Policy Research. Dr. Duncan received her B.S. in physical therapy from Columbia University and her Ph.D. in epidemiology from the University of

00065

North Carolina. She has published extensively on the subject of functional outcomes and their measurements in stroke. Her topic today is selecting outcomes to inform policy.

DR. DUNCAN: Thank you, Dr. Miller, and good morning. I do have some conflicts of interest to share with you. I am the principal investigator of an NIH-funded study funded through the National Institute of Neurological Diseases and Stroke. It's a randomized clinical trial Phase III of a walking recovery intervention called the LEAPS trial, and that is currently ongoing. I'm also a consultant with Glaxo-SmithKline, I'm a paid consultant to design a study to evaluate a drug to promote neurogeneration and plasticity. And I also am a consultant for Bioness to design a study to evaluate the effectiveness of a functional stimulation orthotic. I also need to say that I have spent well over 25 years of a career evaluating outcome measures in stroke.

The purpose of my presentation this morning is not to give you the specifics of all the elements of outcome assessment, but I was asked by CMS to come today to speak from a broader perspective to give you some conceptualizations of how we should

00066

go about selecting outcome measures. As has already been mentioned this morning, selecting outcome measures and designing trials in stroke rehabilitation is extremely complex. As Dr. Miller and our Canadian colleagues have suggested, we have to consider the various etiologies, the heterogeneity of the symptoms, the variability in severity, the time since stroke onset, and the possibility of spontaneous recovery. So I'd like to give you a concrete example about time since stroke onset. We know after years of evaluating stroke recovery that the most dramatic recovery following stroke occurs in the first month and the trajectory recovery continues for three to six months. The trajectory of that recovery varies by severity and in fact those individuals who have mild strokes may achieve their functional independence in activities of daily living by three months, whereas more severe strokes may have a much longer trajectory of recovery. As I will point out to you later, not only in selecting outcome measures, the time since stroke onset is extremely important. We also must consider variability and severity in assessing this very heterogeneous population with heterogeneous symptoms and

00067

heterogeneous severity, that one outcome measure does not fit all. For example, it may be appropriate, highly appropriate to select basic activities of daily living as an outcome measure for the more severe stroke patients. However, individuals who have mild to moderate stroke still remain with significant residual deficits, and the ADL measures may have a ceiling effect.

So as Dr. Miller mentioned and I've modified a little bit for simplicity today, the over-arching model that drives us in selection of outcome measures is the ICF model, which includes body function and structure which I've chosen to label as impairment, activity, participation. All of these factors are modified by the health condition, the disorder of the disease, and the contextual factors with which the patient functions.

Now I want to make a very important point as we select outcome measures in stroke. We're also selecting outcome measures in a population that is usually not healthy, and there are a lot of competing comorbidities. Stroke does not usually happen to a healthy brain. So when we look at health conditions, we also have to consider competing comorbidities across the course of time.

00068

So I could give you many, many examples of measures, and to be quite honest, one of the major problems that we have in rehabilitation research and in stroke research is we've got too many measures. As already has been pointed out and very disappointingly, in a review of the clinical trials there were 45 measures of ambulation. That is not necessary.

So let us just highlight a few measures that are commonly used. In looking at body functions and structures, impairments, we have the Fugl-Meyer motor/sensory assessment, which is the most commonly used measure to test motor recovery in all randomized clinical trials. We have a very standardized measure of balance called the Berg Balance Scale. We have the MMSE which doesn't function very well in this population but is commonly used. And we have other cognitive assessments like trail-making or digit symbols from the WAIS.

And then we have the scope of activity measures and I've used gait velocity as an example, six-minute distance, step activity monitoring, the functional independence measures, and instrumental activities of daily living.

And then participation really deals with

00069

what role functions you have. Now we will use that model to guide us as we select outcome measures, and I will come back to that model in a moment as we talk about gait and walking as a concrete example. The other thing that we need to consider is are we really doing efficacy trials or effectiveness trials? And as most of you know in this room, efficacy trials really are designed to optimize the chance of detecting a biological effect, that you select few patients under the ideal circumstances. Most often the primary import may be the impairments that the treatment is attempting to minimize. However, many efficacy trials do include other measures of activities and quality of life. But in an efficacy trial in which you're trying to show the biological plausibility of this intervention, a primary endpoint may be an impairment level measure. An example of this, if an intervention goal is to improve motor control of the upper extremity, you may use grip strength as an appropriate outcome measure for an efficacy trial. Effectiveness trials, on the other hand, determine whether the interventions have beneficial results when they're administered in the context of ordinary clinical practice. The studies are broadly

00070

conceptualized and they should use heterogeneous samples. The outcomes in effectiveness trials should be relevant to health and function.

Now efficacy trials are necessary and often prerequisites to effectiveness. And a major problem in rehabilitation is that we have not moved, however, efficacy trials to effectiveness studies to inform policy, and most all of our rehabilitation technologies are assessed for efficacy rather than effectiveness. But it's effectiveness trials that we need to inform policy and when, in effectiveness trials we need and we should measure impairments, but the impairments must be related to changes in function and disability to inform policy.

So what are policy-relevant measures? Policy-relevant measures are clinically relevant outcomes of substantial health importance. They must be ecologically balanced indicators of population health and function, and as already mentioned several times this morning, they must be reflective of sustainable outcomes, not simply outcomes at the end of the intervention.

So to inform policy decisions, clinically functional and social relevance, measures that include this range improve activities of daily

00071

living. We all value the ability to be able to take care of our bodily needs and to be able to move. We value the importance of mobility in our life and safety in mobility, and we also value things like using your hands. We also value instrumental activities of daily living, and instrumental activities of daily living are things such as can you take your medicine, can you prepare a meal, can you balance your checkbook. These activities are extremely important for you to be able to accomplish these to live in the community independently. So these are what I call no-duh outcomes, they have clinical, functional and social relevance.

Now another important factor that we should consider is shifts in disability states and in fact if you think of it, probably the most successful trial in stroke was the NINDS trial of TPA, and that trial was based on the shift in disability states using the Rankin scale. So shifts in disability states means that we move from levels of dependence or independence using global measures like the Rankin scale, or simple measures such as can you walk at home or in the community. Those are socially, clinical and policy relevant outcomes.

Now here's an example of what I call the

00072

Rorschach test of is this is a good outcome. This is not a stroke survivor, it's actually a survivor of a spinal cord injury, and this was his walking capability when he came into a trial and this was his walking capability afterwards. It's just one picture of one man and I think we would all agree from this test that it's quite obvious from the patient's perspective and the clinical perspective that this is a relevant outcome. But we don't have the chance in large randomized clinical trials to take individual snapshots.

So when we think of defining definitions and shifts in disability, we need to be very specific. And I want to use an example in walking recovery. Now as we heard this morning, there are 45 measures of ambulation in randomized clinical trials of stroke. Unacceptable. In reality, what do we really want to know in walking recovery? Can you walk or can't you walk? How fast do you walk? What is your endurance for walking? And do you walk in your usual daily activities? That's the scope of what's clinically meaningful in walking recovery.

And I want to use gait speed as an example, and Dr. Studenski in her next presentation will expand upon this concept.

00073

Gait speed reflects the functional and physiological changes post stroke. It's related to the severity of impairment in the home and in the community, and as Dr. Studenski will point out, it's a predictor of health status and functional abilities. We know very well that for example, if someone walks less than .4 meters per second they are limited to household mobility. If they walk between .4 to .8 meters per second there may be limited community ambulation but they're not independent. And greater than .8 meters per second, they can walk independently in the community. We can identify a state that's meaningful to those particular parameters of gait velocity.

Now we can also look at severity. We can't assume because of the heterogeneity of the severity that we can necessarily have one metric of success, and I'm going to use gait speed and walking as an example. For example, if someone has a very severe stroke, the probability unless we find the cure for stroke is that they're not going to become fast ambulators and they may not become independent community ambulators, but they may become independent in their home, and again, that has clinical significance.

00074

So if someone walks less than .4 meters per second you want to know, could they transition to a state that they could walk greater than .4 meters per second. Or if someone walked .4 meters per second but less than .8 meters per second, could they transition to community ambulation.

Now what we did in a prior randomized clinical trial is we defined successful walking using this sliding dichotomy and we found that if individuals, either the severe individuals who walked less than .4 meters per second, after the intervention who walked greater than .4 meters per second, or if you walked greater than .4 but less than .8, could you now walk greater than .8 meters per second, was that transition in walking ability relevant to anything else? And what we demonstrated, that those individuals who made those transitions had improvements in self-reported ADL, IADL, 77 compared to 69 if you were a failure in that transition, 77.6 to 65.5 for mobility, and they also reported improvements in quality of life in their role functions, both emotional and physical role functioning. So this shows a meaningful transition in a disability state and in this case it was walking.

00075

So gait speed predicts hospitalization and improvement or decline in gait speed predicts morbidity and mortality, as Dr. Studenski will demonstrate in the next presentation.

But let's go back for a moment. Transitions in disabilities are very important outcomes, but when do impairments become meaningful outcomes to inform policy? Impairments become meaningful if there are established risks of bad outcomes, and the best example of an impairment is a swallowing dysfunction, because we know that if someone has a swallowing impairment it can cause aspiration which can also be fatal. So if you have a technology that will influence swallowing, then that impairment level measure may be highly significant to inform policy.

Impairments are also useful outcomes if they are very what we call distasteful symptoms. In other words, both personally and societally we do not accept that individuals live in pain, so pain is a body structure and function outcome, and if you have an intervention that controls pain, that impairment may be an appropriate outcome.

Now the other scope, which is rarely done in rehabilitation trials, is that outcomes in which

00076

we can demonstrate that we've reduced important complications is very, very significant, and the best examples come from falls. Falls contribute to morbidity, institutionalization and mortality. Falls are common, even among those who may be independent in activities of daily living, and falls are very important, relationship with fractures, high mortality in the elderly.

Falls are common in stroke and simply they're bad. 73 percent incidence of falls post stroke, and these are individuals who have returned to the community living, a fourfold increase in falls risk. Of those who fall, stroke survivors experience a tenfold increase in hip fracture compared to non-stroke, and limited mobility leads to social isolation and depression. So falls are important.

Now I just want to share with you a study that we did do with a colleague, Heather Whitson, in which we looked at a cohort of elderly male veterans. And we looked at individuals who came into the VA system with a diagnosis of stroke and we looked across time for two years to see what was their incident fracture rate. And we looked in the group that had FRGs four to seven. And what we found, for two years there was a 4.7 percent incidence of

00077

fractures in that population, and we also found that there was a difference in fracture risk based on your functional independence measure scores. Simply stated, those individuals who have very limited ADL capacity are not mobile enough to fall and those who are highly recovered can deal with the stresses of the environment and don't fall and fracture as frequently, but it's the moderate group of individuals who have the highest risk of fractures. So it's not a linear relationship, it's actually curvilinear. Again, this is another example about why you need to consider the severity of the population as you select the outcomes.

Now I want to share with you some results from our ongoing trial. I have no idea, I am blinded, my assessors are blinded to the outcomes of this study, and this trial is currently in enrollment. But as of a few weeks ago we had 201 individuals enrolled in this trial. To be enrolled in this walking recovery trial you must be living in the community and the individuals have a Rankin score between two and four, moderate levels of stroke disability.

Among 201 individuals, these are incident falls and fractures, so these 201 individuals, only

00078

79 have been followed to a year. There are 241 falls among 89 individuals, and we have 11 fractures. And within the last couple of weeks, we have actually had two more, and this is a high fracture event rate. Now to be quite honest with you, in this trial funded by NIH, falls is not the primary outcome, actually gait velocity is. I can assure you, though, if there's a difference in our groups in the fracture rate, that will be very important and will probably have major influence with policy. So that's what I call a no-duh factor. Reduction in falls is a primary outcome with tremendous public health significance, and it is not a rare event in stroke patients. Rarely, rarely, rarely do any of the studies attract such outcomes with substantial follow-up to have an impact. Now, there are challenges of using only ADL measures. If you have a mild stroke the probability of you becoming independent in activities of daily living at three months is 90 percent. However, individuals who have mild strokes continue to have residual significant disabilities that may be impacted by certain interventions. And in fact moderate strokes, as I've just demonstrated, may be independent in ADLs and living at home but have

00079

complications such as falls and fractures. So ADL measures have a very high ceiling effect and again, I have spent a significant amount of my career demonstrating that in the population of stroke survivors, individuals may be independent in ADL but there may remain significant disabilities that affect function, and possibilities of future decline or complications is an important take-home message.

Now an example. One may be independent in ADL but not have any functional use of their upper extremity. Someone can score 95 or 100 on the Barthel ADL index and have no functional use of their dominant upper extremity. So we may need to use domain-specific assessments in some cases. So interventional studies, for example for upper extremity recovery, may use very specific measures that capture upper extremity use, and the ADL measures as the Barthel or the SAM are simply not adequate.

When giving an example from an article published in JAMA, the Effect of Constraint-Induced Movement for Upper Extremity, published by Steve Wolf and colleagues, I won't get into the specifics of this design. In my opinion it didn't have the right

00080

comparator group but the outcome measure selection was superb. So in this particular study of constraint-induced movement for the upper extremity, they selected measures across all the domains of the ICF. They looked at the measure of motor control using the Wolf motor function test, which would be a body structure and function measure. They looked at use of the upper extremity, could you functionally use the upper extremity as reported by the motor activity log. And they also looked at the patient self-report of difficulty using the stroke impact scale hand function measurement.

And what they demonstrated is that they, with this CIMT intervention, improved motor control, they improved use of the upper extremity, and they improved the patient's reported ease of using the upper extremity. In other words, their selection of outcome measures told a story, and this is a quote specifically from their article: "The paretic upper extremity was used at least half as much as before the stroke on twice as many activities following the interventions, and that this behavior persisted through the 12-month follow-up." That is a convincing story for constraint-induced measurements from an outcome perspective. I can't discuss, or am

00081

not going to discuss the comparator model, but from an outcome perspective they demonstrated it very effectively.

Now there are many other issues to consider in outcome measurements to inform policy, and I want to leave you with at least two things that you absolutely cannot, you cannot avoid in these study designs, severity and time post stroke.

The other factor, and I think it follows up on some of the questions that were asked before, is that as you select outcomes and design your studies you have to consider the exclusion or inclusion criteria. In reality, most of the studies that have been done so far in rehab and recovery have very restrictive inclusion-exclusion criteria, and stroke is a very broad condition with tremendous competing comorbidities, and the number of subjects enrolled are too few to be generalizable to many and most of the patients we see in the Medicare population.

I'm just using an example of a Meta-Analysis of Therapeutic Effect of Functional and Transcutaneous Electrical Stimulation on Improving Gait Speed Post Stroke, and this was an article published in the Archives of Physical Medicine Rehab.

00082

There were only eight articles that met their criteria for meta-analysis. The number of subjects enrolled ranged from 13 to 32, and given the heterogeneity of comorbidities, symptoms and severity, you bet you didn't capture those in 32 patients. The stroke onset was chronic, it ranged from 12 to 51 months.

As our Canadian colleagues mentioned, this is not uniformly defined, definitions of chronic or subacute or acute. Overwhelmingly, though, most of the studies done in stroke are done in the chronic population greater than six months post stroke, and that's for a lot of reasons, ease of recruitment and stability in recovery, but it doesn't address the effectiveness of this intervention, or even the most appropriate outcomes in the early stages.

And then we had variability in baseline gait speed across the studies which ranged from .19 to .88 meters per second. .19 meters per second, for those of you who are not familiar with the ranges of normal gait velocity, is extremely impaired and barely mobile. And .88 meters per second you can walk in the community, it approaches normal ambulation speed for an elderly population.

So in measures to inform policy you must

00083

have functional and health relevance. We have to evaluate the sustainability of the benefits and we have to describe more specifically the characteristics of the subjects who would benefit and we have to have large enough samples not only to power our studies, but also to generalize our results to a broader population.

And with that, I will conclude and thank you. I will take any questions. Did you have a question?

DR. DANIS: I really appreciate the perspective you have, it seems very appropriate. I wanted to just ask, though, if you have the clinically meaningful outcome measures such as falls being nonlinearly related to the more functional level, how do we begin to make inferences? You know, you're going to have to measure so many things. I wanted to ask that and also whether because of the variety of severity, could we design studies in a way that allowed for doing some variable use of measures so that you have more stringent tests used in your least disabled group and just work your way up or down.

DR. DUNCAN: Well, let me take the second question first. Absolutely. We have to use sliding

00084

dichotomies of outcomes because individuals, and in fact in my own study which I designed, we have defined success with two definitions depending on the level of severity of walking speed on randomization in the trial. And so we have to employ those types of shifts in disability states, similar to as they did in the TPA study from NINDS, that you have a shift in disability state of zero and one, which you've cured, or you could have a shift in disability state from a four or five to a three, which is again highly clinically relevant. So we do need to consider different definitions of success and you can do that in a trial design by defining that a priori. So yes, that has to be considered.

We also need to understand the scope of deficits that occur after stroke. I know that we have an impression that most stroke patients are severely impaired waiting to go to nursing homes. That is actually not the picture of stroke survival. The majority of stroke patients go home, they live independently but they live with major sequelae, vascular cognitive deficits, limited mobility, and we've published those effects. So you have to consider the whole range.

DR. DANIS: And what about the lack of

00085

linearity and the relationship?
DR. DUNCAN: Well, again, you have to go back to the idea of severity. And falls and fractures is a geriatric condition, not just a condition of stroke. We know that a third of all individuals who are over the age of 65 fall in a year, and we've demonstrated from geriatric research that you can reduce the risk of falls with very specific interventions. So what it will require are larger sample sizes, and as Dr. Studenski will point out, different methods of analysis and more a survival analysis and that type of thing. It will not be answered by small Ns.
DR. SATYA-MURTI: We have -- yes, Ms. Richner. After that we should close it, because we have an opportunity for afternoon questions.
MS. RICHNER: A quick question. We heard from Dr. Miller at the beginning, and one of the questions that CMS is grappling with again is the clinically meaningful results within a drug, device or intervention, and to me when I was looking at even the HTA evaluation that you did, it includes everything from Chinese acupuncture to TPA, which you mentioned before. It seems to me that CMS needs to have some idea about how to look at the acute and

00086

immediate instruments and metrics that are being used there, versus the neurorehab interventions that are currently on, so it seems to be an apples-to-oranges comparison. Is there some way to look at this differently to take, because I think it's just --

DR. DUNCAN: I'm sorry, I can't hear you.

MS. RICHNER: It just seems to me we have to be able to help CMS to say what are the interventions, acute intervention, what are those outcome measures that are meaningful for that drug device kind of thing, versus those sort of longer-term interventions.

DR. DUNCAN: Is that a comment or a question?

MS. RICHNER: It's a question, how do we do this?

DR. DUNCAN: Well, I believe that first of all we have to reduce the number of measures that we begin to accept. That's not saying that, for example in walking recovery, 45 measures is not an acceptable battery, given that we know what clinical relevance is for walking, right? So the field has been very profuse in developing new measures but without really selecting the most clinically relevant measures, so we have to establish clinical relevance and we have

00087

to narrow our scope. That being said, we also have to make the measures consistent with our intervention. To be just a no-duh, you wouldn't take a gait velocity measure for an upper extremity recovery. And basically what, that's also what we've done many times with ADL measures, we've taken ADL measures that are the most basic functions that we all value, but if they're the only things that we can do, we wouldn't be too happy, right? So we have to be very domain-specific in some cases.

DR. MILLER: Dr. Duncan, thank you very much. Please let me now introduce Stephanie Studenski. Dr. Studenski received her nursing and medical degrees from the University of Kansas and a master's in public health from the University of North Carolina. Her post-doctoral training includes fellowships in rheumatology and geriatrics at Duke University Medical Center. I calculated that you didn't sleep for ten years or so, is that about right?

DR. STUDENSKI: Yes.

DR. MILLER: Currently she is a professor of medicine in geriatrics at the University of Pittsburgh where she is also the director of clinical

00088

research at the university's institute of aging. Dr. Studenski is well known for her research that studies the consequences and effect of interventions surrounding balance disorders in older adults. She has also published on the topic of clinical performance measurement and the concept of informative and meaningful change in that domain. Her subject matter today is what is a meaningful benefit in terms of health policy.

DR. STUDENSKI: Good morning, and thank you, Dr. Miller. Before we go on, I want to hope at the end someone in the audience comments on what's happening in other places with measurement, like the NIH toolbox, and the major effort to get rid of floors and ceilings with the item response theory and the new expanded measurement strategies. So that hasn't been touched on, it's not part of my talk, but it may be very relevant.

So again, more to my talk here today, I was asked to address my disclosures. I have no conflicts of interest with device companies. However, I do consult regarding measurement of function with multiple pharmaceutical companies, including Merck, Glaxo, Pfizer, Lily, and Asuvio, and I do have NIH and VA funding largely for work related

00089

to disability and function.
So what I'm going to address today is trying to think our way through, how would you decide how much benefit you need to achieve to be useful and how do we even begin to try to measure those sorts of things, and then we'll talk about what are some of the challenges in the field.
So we've heard today already, a statistician cannot tell you what is important, they can only tell you if it was likely to occur by chance. A P value of .001 tells you nothing about whether you want to reimburse or pay for a service, largely because if it's a very large study, it could be a very small, small effect. And as we've heard in stroke rehabilitation, you often have small studies where a potentially very important effect might not be statistically significant but still be clinically very important. So the bottom line is, the clinical significance or patient benefit is a value state and it is informed by patients, families, providers, they're the ones who tell us what's important. But what we can do as researchers and reporters and interpreters of evidence is have a good understanding of how to present that information in ways that make it more interpretable.

00090

So what do we mean by a benefit? Well, it could be that the patient's better or it might be that we kept them from getting worse or kept something bad from happening. So what we have to do is say well, what do we mean by better or worse. We've been through this, I'm not going to do this in detail, but obviously there's many different things we can measure based on our conceptual frameworks, and we can certainly also be interested in events and states, and have in mind who's telling us what and how it's measured. Then we are struggling with this idea of what makes something objective and how do we capture these nice psychometric properties. Another issue I think we need to be thinking about more carefully is when to measure. We've heard some things about sustainable benefit, but I also think that stroke and many disabling processes don't change in a linear fashion, they fluctuate, people have good and bad days, good and bad weeks. So we often used fixed time points and say what's the effect at three months, but you know, there might be studies where something that's really important might be timed to first event, how long did it take after the stroke until someone achieved a threshold of independent home ambulation, and maybe

00091

the intervention gets you their faster.
I'm also very interested in low demand but frequent monitoring, and I will be talking more about that as we move on. An example would be accumulating caregiver hours over time.
We've heard a lot about psychometric properties and I just want to emphasize that over the years I think the psychometricians have taught me a lot, but I also think the clinicians and the patients have taught me a lot, and face validity, which is, does it make sense, does it sound like it's measuring what it's measuring, is a very valuable element of psychometrics. So you can see all kinds of numbers jumping around with statistical properties, but face validity is very very important.
So let's talk about this idea of developing a criteria for what's better or worse. You can develop these kinds of ideas for continuous measures where you have whole scales or sets of performance results, and you can also develop it for categorical measures and I think I'm going to tell you very briefly. There's a huge world of literature, people who spend all their time thinking about these things, but the bottom line for me is, anchor-based methods are best for face validity

00092

because what you do is you say to the patient or the family or the provider, is this person better overall? And then you go back and use that as an anchor to say how much do they have to change in this scale or that score, or that gait speed to be detectable compared to the people that were not reported as having changed.

We heard just briefly about this idea about minimally important change. The concept there is where can you get enough signal-to-noise ratio that you can actually hear the signal. And I think it's an important number, but I want to emphasize that it's not the only amount of change I care about. I think substantial change, or changed a lot might be a really important state, and I'm not sure that my goal with treatment is the least detectable change, maybe it's the amount of change a person thinks is really valuable, so we often use anchors in a variety of ways to understand how much change is important to people.

There's a variety of statistical methods based on a lot of math that are overall called distribution-based methods that you can use. They are the best for precision and so you will get the most tight confidence intervals when you use these

00093

sorts of measures, but they're dependent on what sample you calculate them on and they're not necessarily linked to values, they're just calculations.

So I like to use both, and what I'm interested in is if you went through a whole series of approaches to what's important to patients and what is a nice psychometrically precise reliable measure. You can try to see if you can come up with some consistency. So I don't know if I have a pointer here. On the top row is, let's see, so the first column is just talking about three different measures, gait speed, six-minute walk and the short physical performance battery, which is a combination of walking, chair rises and balance tasks that's used a lot in geriatrics. Across the top row are the kinds of things we can measure using these distribution and anchor-based methods. And in the far right column you can see sort of the summary that if you look across all the different ways you calculate these things, you can come up with some summary indicators that are reasonably consistent across all of these measures.

And as you can see for example under gait speed, we're interested in both this minimally

00094

significant as well as a larger magnitude of change. And I think these kinds of estimates can be useful, for example, if you wanted to evaluate a treatment for policy reasons and you could determine what proportion of your intervention subjects achieved a substantial change, not just a minimal change. The other kinds of things that you can do, and I won't do this in detail, is you can go back and say okay, having calculated these minimal and substantial changes, what kind of impact do they have on other things that are going on with the patient at that time. And what we were doing here is taking data from a large clinical trial, the LIFE study which was 424 older adults receiving walking and strength training, and you can look at both decline and improvement in performance measures and the impact it has on a whole variety of health and function measures. So that is a sense of sort of some concurrent validation here, and the magnitude of how much does a performance change affect people's perception of their health and function. Here's something else we did. We're interested in predictive validity, and what you see at the top is a line that is green for a short distance and then blue for a long distance. So this

00095

is a study I did ten years ago. We were measuring people in their homes every three months, about 500 older adults, and then we have been following their survival ever since. So our question here was, what is the effect of short-term change in function on long-term survival? So what we defined was people who improved over that first year in any one of the set of measures listed on your left there, gait speed, short physical performance, SF-36. We used these substantial change measures that we've calculated and we said okay, did you ever achieve this criteria for improvement during that year, yes or no, and then what happened to your survival. And what was striking to us is that out of all these measures, the only one that predicted nine-year survival were people who improved in gait speed over one year had substantially better nine-year survival than people who didn't. And this is what it looks like as a survival curve. Overall, the people whose gait speed ever improved .1 meter a second during that year, 30 percent died over the next nine years, and the people who never improved that much, 50 percent died over the next nine years. And we did subgroup analyses and this was true for age groups, different walking speed, different

00096

comorbidities, gender, ethnicity, health status, functional status, et cetera. Dr. Duncan spoke a little bit about thresholds. I've done some work in this area. These numbers are a little different than hers since I work largely with diverse community dwelling older populations with moderate amounts of disability, and so it may be that these thresholds are different in stroke populations than they are in other populations. There are certainly reasons to think that that could be true.

Another question that I spent some time thinking about is how do we decide, we have a treatment group, we have a comparison group, what's an important difference between the two groups in these outcomes we've talked about here? I think the hardest way to interpret these numbers is giving two means. So I don't care if we're talking about FIMs or Fugl-Meyers or whatever, I give you a mean of one group and a mean in another group, and you tell me how important is that difference. That's a tough way to understand it.

I think rates are a little easier to interpret but we might still need informants, patients, clinicians, families, policy-makers to say

00097

how much is worth it. One of the most useful ways to do this is to use a widely valued approach which is called the number needed to treat, which many of you may be familiar with. I think I'm going to show just briefly, the idea is that you look at a rate difference between two arms. And so you had, 70 percent of the intervention group had a gait speed gain and 40 percent of the control group did, so the difference between the two rates is 30 percent, and you invert that and you end up saying well, that means that you would have to treat three-and-a-third people to get one who benefitted specifically from getting this intervention. In this way of doing things you can compare between treatments, how many people would have to receive the treatment in order to benefit.

It's virtually never true that none of the comparison group gets better, so there's always this idea that there needs to be addition of benefit. So it would be up to the policy-makers, the providers, the patients to say well, what would they be willing -- are they willing to have five people be treated to have one benefit? If it's a very serious outcome like that, we provide treatments right up to treat where the number needed to treat is 100, 200,

00098

500 to prevent the death if it's young persons and heart attacks or motor vehicle deaths. I don't know if you can see this, but this is this idea of looking at time to event as a rate-related thing that might be more interpretable to policy makers and families. So I'm just proposing that what if we were looking at recovery of walking ability, that we could perhaps measure every week, have they achieved some level of independent home mobility, and compare two arms and look at these rates of achieving this important outcome, and then ask our patients, families, providers to say how much of a difference would be useful to you. But I think this is a metric they could respond to more easily than being given two mean numbers. This is something I'm very interested in and spending a lot of time with in the last year. I'm going to give you an example of work about time in state as a measure of treatment benefit. I've been doing this work in the area of cancer treatment, not stroke rehabilitation, but I think it applies. The set of table and text on the left is saying what if we had a new treatment whose goal was to prolong survival? The one on the right is what if we had a treatment whose goal was to increase tolerance of

00099

cancer treatment. And what we're seeing on the left is we have a new treatment versus the usual treatment, and the usual way you'd report this on cancer is survival days. And what you would see in the New England Journal of Medicine is that this is a fabulous new treatment for this terribly rapidly fatal cancer, we increased survival from 160 to 200 days, and that would be very important in that world. What I'm saying, if we looked at time in state and we cared about what patients have to say, maybe along the way we've been asking them, how many of those days you were alive did you have to spend over half your time in bed and how many days were you unable to go outside with help, and perhaps we should consider the days that you can get out and around as independent days and useful days. And in this treatment and prolonging survival you can see that the new treatment made survival longer but you spent most of your time in bed and having restricted days, so the usual treatment had a lot more independent days than the new treatment. We don't provide information like this in many of our worlds of trials.

And on the right my goal was to increase tolerance, I'm interested in treatments that are more

00100

gentle for cancer. My survival didn't change at all, but I had a whole bunch more independent days with the new treatment so maybe that would be useful to people. So I'm suggesting perhaps in stroke rehab, because there's all this fluctuation, people have good days, bad days, sometimes their knee hurts, that maybe we would accrue a number of days when you can get out of the house or something like that.

In terms of heterogeneity we've heard a lot about these issues, outcome rates varying. I think I won't go anywhere else with that right now. We have been beginning to explore subsets of people in terms of trying to determine whether these various ways of anchoring and calculating meaningful change might vary with severity of disability, would a smaller change in gait speed perhaps be more of a signal in a very slow walker. You need larger samples to do that and we are doing that with sample sizes in the thousands now.

Again, I think we've addressed some of these issues about the indicator of benefit needs to make sense based on the amount of disability in your stroke population.

We've talked a little about duration of benefit as an issue. You certainly could use number

00101

needed to treat for the sustained benefit, so with how many are still independent in community ambulation a year after treatment, and you could do this rate difference and calculate number needed to treat. I'm very interested in what we might do with time in state over a longer period of time so we could be looking at what's happening with household mobility over a year.

So, my pain points are that we should be thinking about ways of reporting on patient treatment effects that are closely linked to value, have strong measurement characteristics but are more easily interpretable by decision-makers than just group means. Mean values for treatment arm are the hardest to interpret from a clinical and policy point of view and probably should be avoided. Time to event or time in state might be some novel ways that we could account for the fluctuating nature of disability, and all that we can do is provide information in interpretable format. The decision about what's worth it is still a social decision. Thank you.

DR. SATYA-MURTI: Thank you very much. We'll take a 15-minute break for PDL, physiological demands of daily living, and reserve the questions for the afternoon please. Thank you.

00102

(Recess.)
DR. SATYA-MURTI: We had a new panel member join us, Dr. Sloan. Dr. Sloan, would you identify and introduce yourself and mention if you have conflicts of interest, because you couldn't be here earlier?
DR. SLOAN: My name is Andrew Sloan, I'm an associate professor of neurological surgery at University Hospital Case Medical Center and I have no conflicts.
DR. SATYA-MURTI: Thank you. Maria will introduce the speakers next.
MS. ELLIS: Now we'll have the scheduled public speakers. First is Dr. Michael O'Dell, and you will have five minutes.
DR. O'DELL: Good afternoon. Thank you very much for the opportunity to speak with you today. I'm representing the American Academy of Physical Medicine and Rehabilitation. The Academy -- first of all, I have no financial disclosures to offer. I frequently prescribe functional electronic stimulation, robotic and partial weight-bearing strategies, but I have no financial interests in those companies.
Rehabilitation medicine, for those of you

00103

who don't know, is the field of medicine that addresses function which is best defined as performance of individuals. What we do, we do in teams with our colleagues in physical occupational therapy, speech language pathology. We're not a pill, we're not a procedure, we're a process. And as I think you've heard today, that lends part of the difficulty in doing research in the area.

AAPMR is the largest professional organization representing physiatrists or rehabilitation medicine physicians in the country. Our members along with our colleagues in neurology and neuroscience and rehabilitation professional researchers have really been at the forefront of a philosophical and a technological revolution in neurologic rehabilitation. My point of view is as a clinician, I see and I evaluate patients with neurologic disease and stroke every day, and also as a researcher addressing mostly FBS robotics and psychometric properties of scales at the moment.

I wanted to bring out just a few issues related to the methodology of the research in the studies that we're talking about today. Much of what I'm going to talk about has already been mentioned by the previous speakers and I'll be able to go fairly

00104

quickly. There's been really a revolution in the approach to neurologic rehabilitation from a very compensatory strategy, getting folks to do better, to more of a remedial approach, can we actually change the natural history of the motor recovery from stroke. There are very different approaches to the population and as we look at the methodology of the research to study these, I would emphasize the importance of understanding the difference between impairment and activity-based outcome measures and how that plays in to figuring out whether one or the other actually works.

The other issue is very clearly from a motor recovery standpoint, specificity of exercise as mentioned earlier this morning, and particularly repetition of exercise is crucially important. And not just a few repetitions, a lot of repetition, which certainly indicates that the length of treatment may need to be longer than we have thought of in the past.

The explosion of technology available to rehabilitation professionals over the last ten to 15 years is really quite impressive. Functional electrical stimulation, both upper and lower

00105

extremities, robotics the same, upper and lower extremities, and newer developments in virtual reality treatment. TMS and exercise approaches. As Dr. Duncan mentioned earlier, constraint-induced motor therapy as well as some of the partial body weight supporting strategies. One of the areas that we're really only beginning to understand is how to use motor learning theory in what we do on a daily basis regardless of the technology with stroke rehabilitation, and the use of pharmacology. Without -- I just want to emphasize a couple of points in terms of issues about bridging the research and the clinical care, perhaps a little bit different take on speakers previously. I think it's very important for this group to ask the question, can there be a durable treatment effect without durable treatment? And yes, certainly providing an intervention and then looking at what the outcomes down the road might be are crucially important, but we don't expect a limited period of time, treatment with statins and then expect that the cholesterol is going to remain low, and don't treat for a limited time for insulin and expect that the diabetes is going to be cured. So looking at exactly the question to be asked, and is it reasonable that

00106

if we don't provide ongoing treatment whether we're going to see ongoing improvement.
And the other important point, again, understanding the endpoints. Are we measuring motor recovery, the speed of movement, the accuracy of movement, are we measuring the activity important to someone or are we measuring their function in the community? And again, making sure that we know what questions that we're asking and that we know how to measure them. This slide you can look at at your leisure and I will be happy to answer any questions later in the afternoon, but I think most of the points in terms of possible strategies have really been addressed by the speakers earlier today.
So in conclusion, the American Academy of Physical Medicine and Rehabilitation looks forward to working with CMS and other groups in really exploring the best methodology to provide the best treatment and access to that treatment for our patients. It's very clear that repetition and specificity of exercise, whether technologically mediated or not, is going to be a very important area for further research.
I think it's also very important to understand and realize, there are pockets of very

00107

reasonable and credible research in some of these areas already. I would mention certainly upper extremity functional electrical stimulation and upper extremity robotics. There is a good deal of at least reasonable research at an impairment level already. Again, the American Academy of Rehab looks forward to working with CMS as we bring these technologies and the very best rehab care to our patients and our stake holders. Thank you very much.

MS. ELLIS: Dr. Gad Alon.

DR. ALON: I want to thank the committee for the opportunity. I'm Gad Alon, I'm an associate professor at the University of Maryland School of Medicine department of rehabilitation sciences. I'm currently a paid consultant for Bioness but I am not being compensated for my presentation today.

With existing intervention, only 12 percent of stroke survivors are likely to recover full function of the upper extremity. 65 to 70 percent will recover the ability to walk, but at a very slow pace and very limited distance. Many will depend on some assistance, cane or walker, or orthotic device, and at least 25 percent or higher are likely to fall.

The critical question that I ask both as a

00108

researcher and clinician are, what are the physical rehabilitation options to help a stroke survivor? Option one is actually no intervention. This option is strongly contradicted by evidence-based practice and offering it to the patient is in my mind unwarranted and maybe even unethical. Option number two is to provide task-specific oriented exercise training over three to 12 months, and some continued progress can be expected during that period. The difficulty is that most patients cannot practice task-specific exercise because their upper and lower extremities are paralyzed or paretic, and they are unable to activate those muscles appropriately. So option number three is actually to combine task-specific exercise with functional electrical stimulation, or FES, and that's where my area of research has been focusing. But the question then is of all of the therapeutic technologies available today, why FES and not robotic or partial body weight support or some other exercise technologies? Well, my answer is that FES is the only and the least costly technology that is available to date for daily training in the

00109

rehabilitation center, in the outpatient clinic, in the home, and most importantly as a patient home self-administered training option. The fact that the patient can continue to practice on his or her own provides the best chance for further improvement of motor control and functional gain, even in the chronic paralysis or paresis.

Studies provide compelling clinical evidence that early initiation and prolonged application of an electrical stimulation program are reasonable and in fact probably needed. There are many, many studies and obviously we provided, or I provided and the committee has it from many other resources, about the data available today.

But there are obviously major issues related to the outcome measure and selecting the appropriate test is a challenge because there are too many. As Professor Duncan said before, I believe also there are too many tests that have been validated and are highly reliable and reproducible, but are not necessarily relevant to FES, to what FES is expected to improve. For example, the FIM and Barthel indexes are practically nonrelevant to FES.

The most relevant tests for the upper extremity are those that measure the ability to open

00110

the hand, to grasp, to move, and to release objects, and for the lower extremity, those are to measure the ability to walk at certain speed, the distance, and possibly the incidence of fall. Relevant tests must also consider, as previously mentioned, the severity of the paralysis, and consequently my take on all this is that there is unlikely to ever be one test fits all.

In fact, regarding to the FES, I would like the committee to consider that there are actually two options. One, improving function while using the FES, and second, improving function after a period of training with the FES but testing the function of interest without the FES. And when we consider the research option, we need to consider those two options as well in terms of the design.

Because of time I'm going to skip on many of the other slides and I just want to summarize that in closing, after at least 15 to 20 years of FES clinical trials around the world and the cumulative clinical and statistical favorable outcome, it seems to far exceed the inherent limitation in rehabilitation research. Many experts seem to have reached consensus that effective training should be task-specific, the study design must consider the

00111

severity of the paralysis as well as the time since the onset of stroke, and that the outcome measure must reflect the specificity of the technology or the intervention used. From that perspective I hope the committee will revisit extending on the appropriate design of clinical trials as double blind clinical trials are practically impossible in neural rehabilitation. Thank you.

MS. ELLIS: Robert Mullen.

MR. MULLEN: Good morning. Speech language pathology is a relatively low tech field, so no Power Point to present in our five minutes this morning. But first of all, I would say that my name is Rob Mullen, I'm the director of the National Center for Evidence-Based Practice in Communication Disorders at the American Speech Language Hearing Association, or ASLHA. Beyond my involvement with ASLHA I have no financial or other conflicts of interest to disclose.

ASLHA is the professional society in the U.S. for speech language pathologists and audiologists, so we represent in excess of 130,000 members who are clinicians, administrators, researchers and faculty, and we bring to today's discussion a number of actually fairly grave concerns

00112

about today's meeting, which some of you may have noted if you read the written remarks that we've submitted.

We submitted a number of comments regarding the individual questions which you all will be discussing later today, so I'd like to confine my remarks this morning to some of the more global concerns that we have. And one of the primary concerns is that there are no speech language pathologists on this panel. It's also apparent to us that there are no occupational therapists, there are no physical therapists, there are no neuropsychologists. And that worries us, that those huge stakeholders would be excluded from the panel. I thought that Dr. Duncan and Dr. Studenski gave us some very important insights related to physical therapy this morning, and it would have been great, we think, to have folks like that on the panel as well as folks in some of these other disciplines to really capture all of the stakeholders that we feel are appropriate.

In addition to the lack of representation on the panel, we're concerned that there appears to be, or have been at least a lack of consultation as well with these stakeholders. Certainly ASLHA was

00113

not consulted in terms of framing these questions, nor are we aware of any of the other rehab associations being involved. Certainly our sister associations in occupational or physical therapy, as far as I'm aware, they were not consulted either in terms of the development of these questions. I think one of the manifestations of that lack of involvement of these disciplines has to be noted with the development of some of the particular questions. I think you need look no further than question number one to see what we perceive as actually a fairly substantial bias in the way that that question is written. The question refers, starts out by talking about the problems of generalization from study results to large heterogeneous populations and then goes on to raise a question about observational studies, which frankly perplexes us because the notion of generalization is an interpretation issue rather than a study design issue. And so why that leads into the question specifically about observational studies is something that we quite frankly can't understand and it seems to us to be frankly pejorative, and introduces a bias potentially against observational studies. One of the other manifestations of the

00114

concern about the lack of involvement from these major stakeholder groups has to do with the particular outcome measures that are cited within many of these questions, particularly four through seven. We frankly were and continue to be perplexed at the choice of the measures that were specifically mentioned in these questions as they relate to speech language pathology. The measures that are cited here certainly do not reflect current research in speech language pathology, they don't reflect current clinical practice in speech language pathology. Some of them are in fact one of the used measures, some of them basically haven't been used for a decade or more, and there are some very glaring omissions from the list, and we would argue about even the propriety of having such a brief list of outcome measures in the first place. But if there is going to be a list, we really have concerns about how this list was created, we really can't make sense of how that was done.

So, I would like to ask for your consideration in taking a look at the comments that ASLHA has submitted in terms of the individual questions in your discussions later this afternoon, so thank you.

00115

MS. ELLIS: Jennifer French.
MS. FRENCH: My name is Jennifer French, and you do get a second break from Power Point presentations from me, as well as an ease on your back from switching over. Again, my name is Jennifer French, I represent an organization called Neurotech Network, we're a 501(C)(3) public charity. And I do need to state a bit of a conflict of interest. I don't have any direct conflict of interest, but our organization does have about 30 percent of our funding from corporate sponsorships.
In terms of our comments that we would like to make to you today is that we know that stroke is a disabling event and we also know that the disabling events have loss of mobility, cognition, speech, balance and endurance. But there's also a lot of other secondary health considerations that you need to take account.
Neurotech Network, again, we're a nonprofit organization, and we focus on the education of and advocacy for neurotechnology devices for people with impairments. And we believe in the topic of clinical trial design and analysis of neurorehabilitation there are three issues from a patient's perspective that we believe the committee

00116

should consider. First is face validity when you're looking at the comparison group. Second is the gap between the FDA approval and CMS reimbursement. And third is the long-term health care of a stroke survivor.

First when we look at drug trials, the rehabilitation, in terms of rehabilitation is faced with a challenge of defining what the comparison group should be as well as incorporating blinding into the study. In terms of rehabilitation, the patient has to be actively involved in the rehabilitation. Whether it's electrical stimulation of a muscle, gait training with treadmills, or rehabilitation using robotic-assisted devices, the patient is involved in the treatment actively. Therein lies the challenge of rehabilitation of clinical trials. A clinical trial design, we recommend to use a controlled group as those receiving conventional rehabilitation established at the time of the trial design. This will help overcome the challenge by allowing recruitment of a control group from a realistic setting that patients experience in standard of care.

Secondly, I know that part of the discussion is going to be in terms of gaps and

00117

barriers, and our second point is in terms of understanding that the FDA approval process is different from the CMS approval process. It's really not clear to the patient. It is difficult for us to understand why a treatment can be FDA-approved but not covered by CMS. The time frame between approval can be months and even years. In clinical trial design the FDA has a pre-IDE process; if the CMS has a similar process, it's not well known. If there is such a process that exists, we recommend that it have, you have an inter-agency collaboration to aid in the early design of clinical trials in an effort to reduce the gap between FDA approval and CMS reimbursement review.

Finally, a topic that is very near and dear to my heart is the long-term care of the patient and the economic impact. The effect of stroke does not just impact the stroke survivor but also the social network. For instance, if a member of a household has a stroke, another member of the household must become the caregiver. If there's not a caregiver then either one is hired or they are brought into a skilled nursing facility. In the case where a person in the household becomes a caregiver, there's a true economic impact. That person may no

00118

longer be able to work and have to stay at home to take care of that stroke survivor. Now that economic impact may not be felt by CMS, but it's definitely felt by other social agencies and we need to be aware of that.

Also, in addition to daily care giving, treatment of stroke survivors is not autonomous. Treatment and therapy for stroke is not a short-term endeavor, it's a long-term rehabilitation process.

As new treatments are considered, they need to be viewed as a complement to the overall care and not just a stand-alone treatment. This long-term view can help to understand how treatment being tested can impact the care, the cost and the quality of life of a stroke survivor. Ultimately a short-term investment in rehabilitation of a stroke survivor can convert to long-term savings of the overall healthcare costs of that person, and not only improving function but reducing secondary complications, maintaining independence and improving quality of life, not only for the survivor, but the social network and the caregiver. Thank you for your time.

MS. ELLIS: Mary Wagner.

MS. WAGNER: Good afternoon. I'm Mary

00119

Wagner and I have no disclosures today. I want to thank you for the opportunity to speak. I am a speech language pathologist and I am speaking today on behalf of NARA, the National Association of Rehab Providers and Agencies. NARA is a professional association who for 30 years has focused on the business side of rehabilitation. We represent thousands of therapists and 70 business organizations throughout most of the states in the United States. NARA's members are owners or those who manage Medicare-certified rehabilitation agencies, long-term care facilities, certified home health or comprehensive outpatient rehabilitation facilities. NARA's members provide services through physical therapists, occupational therapists and speech language pathologists.

We recognize the importance of scientific evidence and the need for evidence-based approaches to therapy and the need for good solid research, and achieving that goal we recognize has many challenges. Having quality researched evidence to verify therapeutic approaches is a longstanding challenge for the rehabilitation industry. Historically most rehabilitation therapy clinical research comes from teaching institutions or the VA, and it's very costly

00120

for the average clinical setting to dedicate staff that will obtain funding, manage and participate in a clinical research project. As a result, some innovative therapy approaches to therapy may never be shared.

The challenges are well explained in several research articles. Dr. Weinstein from the University of Southern California and Dr. Ludwig from Rancho Los Amigos National Rehabilitation Center and University of Southern California sort of put it in a nutshell. To quote them, the research design is dependent upon internal and external validity needs; ethical considerations, should we provide therapy, this new technique to this population and not to that population; the feasibility and pragmatic concerns and perspectives of the research funders, third-party payers, reviewers, investigators, clinicians, and of course our patients.

Occupational therapists have found there is really not a lot of evidence for the efficacy of specific interventions. One way that perhaps we can look at research is sort of a back door approach, if you will. Look at outcomes, everyone's working on how to come up with looking at outcomes and paying for performance, and maybe if we look at the outcomes

00121

and then go backwards from that, what interactions and what techniques, therapeutic interventions help to make the best outcomes, and then drill down to see what was done in those particular areas.

And then that being said, it's important to keep in mind that a cookbook therapy approach, one where treatment for a defined diagnosis is one treatment is best for everyone, that isn't what therapy's all about. It's the training and skill of the individual clinician that enables him or her to explore diagnostically how a patient learns along with their strengths, weaknesses, comorbidities. Their personality even will help to determine what approach will be most effective in providing positive outcomes with that individual.

However CMS decides to proceed on this important issue, NARA would like to be a bridge between the clinical and the research. As Dr. Studenski talked about, the anchor. The anchor can look at the outcomes and then drill down from there. NARA being representative of that critical connection of the clinical and subsequent business aspects, we would like to be part of the process and we would like to be included as was mentioned by the ASLHA representative.

00122

And when we're trying to look at something that's so critically important, all of the players need to be at the table. We need to have research that's meaningful, makes sense and will work in the real business world, and that's where NARA would like to have a role and be a part of the decisions that are being made as we go forward to decide what are the best evidence-based practices to incorporate for the future of rehabilitation services. Thank you.

MS. ELLIS: Now we'll have open public comments. We have Dr. Mark Pilley.

DR. PILLEY: I understand I have two minutes. And as Dr. Jacques and Murti understand, that's tough for me to do. Mark Pilley, previous contract medical director for Mutual of Omaha for a few years, and then IntegraGuard. I just got finished doing the durable medical (inaudible) for jurisdiction D. Today I'm a consultant working in here representing RS Medical, so I have to disclose a couple of things that are a conflict. I am getting paid for being here today, but my comment is a general comment.

I'm also a fellow and on the board of the American Academy of Disability Evaluating Physicians, and one of the things that struck me with this

00123

particular presentation was the application of the ICF, which I think the AMA guides have just gone to in using that in making a determination of permanent impairment, which might provide an opportunity to provide sort of a global way of assessing impairment of the whole person, taking multiple systems into consideration when calculating that particular impairment rating.

That having been said, the academy of course is a nonprofit academy, I think we just acquired a 503(C), but I don't know that there's many funds in that because we like to do more in terms of research and clinical studies and trials.

But one of the things that we do teach is that pain is a significant impairment and a barrier to recovery, because activity obviously begets activity. One of the things I didn't see presented here was a way of determining improvement in terms of reduction in pain, because in reducing pain, people get up and they do more things. But it also means they're not taking medications that can impair their functionality and in particular narcotics, and I think that is a significant impairment and risk to the beneficiary or to the patient, in terms of the more narcotics you're taking, of course the increased

00124

risk you have of decreased cognitive thinking and of awareness and of falls. So I just wanted to make that particular comment. But regardless of how it's accomplished, I think it's most important to consider that as an inclusion in positive outcomes.

DR. SATYA-MURTI: We don't have any other scheduled or ad hoc speakers. I think we will break for lunch and come back, we're scheduled to come back at 12:35, maybe we can come back at 12:25 instead, or 12:30. Thank you.

(Recess.)

DR. JACQUES: Good afternoon and welcome back. Before we actually resume the agenda where we left off, I just wanted to respond to a couple of comments that people had made. One, I think it's important to keep in mind that this meeting is not about a particular technology, nor is it about a particular modality. And we realize that the rehabilitation of people with stroke, certainly it involves people from multiple disciplines. The composition of the MedCAC panel is based on, the membership of the panel, which is a public process, there is an annual nomination process and if there are organizations that would like to nominate one or more individuals for membership on

00125

the MedCAC panel, certainly it's a public process and one can avail themselves of that. The constitution of the MedCAC panel is based from the MedCAC panel membership.

And again, just as a reminder, neurological rehabilitation is an extraordinarily broad subject. Certainly we could have also talked about spinal cord injuries, we could have talked about congenital problems, we could have talked about all kinds of things, and it would have unfortunately been an unmanageable meeting in terms of size. We chose stroke because of its particular relevance to the Medicare beneficiary population and we recognize that even in that setting, that the conversation may be a little bit narrower than some would prefer. But keeping in mind that we are not making a determination here about the coverage of any particular technology, we do feel that the broad discussion of the methodologic challenges related to determining appropriate outcomes and trial design and things like that can provide some generalizable information that people may find helpful in other settings.

Saty, you want to take it from here?

DR. SATYA-MURTI: Yeah, you've said it.

The idea is not to focus on single treatment

00126

modalities but what kinds of measurements would you like to see, and that will apply to some of the questions too when we get down to it.

I would start off about with my question to the presenters. I would like to confine myself to one question at a time so we get the opportunity for others and not have multiple questions. My question to the two TA presenters this morning would be, we heard about functional electrical stimulation and then when I read your TA I found you had included that, I actually used the search term to go down to see, but you had also mentioned that FES, the sample sizes were small and that you had some questions about FES itself. We heard the benefits of FES this morning, two speakers talked to us and said, but did they satisfy the characteristics of a good study, did they have all the concert requirements, stroke requirements mentioned, or is it too focused a question.

DR. SANTAGUIDA: There were citations that were reviewed in both sections and so the focus was on their methodological quality.

DR. SATYA-MURTI: And how was it?

DR. SANTAGUIDA: I can't recall specifically those studies but we can get back to you

00127

with that information. I'm not sure that they were distinguished among the other therapies.
DR. SATYA-MURTI: So you couldn't find a particularly striking distinction of the FES in comparison to the others that you reviewed is what you're saying?
DR. SANTAGUIDA: We paid no attention in the purposive sampling to what the therapy was. We selected studies based on the outcomes that they evaluated.
DR. SATYA-MURTI: Okay, thank you. All right, one question, so I'll toss it on.
DR. HURWITZ-GERBER: Is Dr. Miller a fair person to ask a question of?
DR. SATYA-MURTI: Sure.
DR. HURWITZ-GERBER: Susan, I really enjoyed your presentation this morning, thank you. I have one question and it's really sort of a clarification. It was your slide which pertained to categories of function within the ICF domain. My sense is that there is interest both from CMS and others who have presented this morning of using this classification scheme in some way or another to either identify opportunities for outcome measurement or for conceptualizing problems of disability through

00128

this model, and towards that end you mentioned that you thought an instrument would have to address the issue of health consequences and functional concerns of patients. I thought that was a really important statement which takes us a little bit further than the standard strict methodologies of outcome measures, of performance, and those sorts. Did I mishear what you said, is this something that you would like us to at least incorporate in our thinking?

DR. MILLER: Yes. That's the answer.

DR. JACQUES: Which question are you saying yes to?

DR. MILLER: All of them. CMS is very interested in the, particularly in the technology field, in the activities domain of the ICF, and particularly in terms of devices is very interested in those measurements or perhaps those categories of function that would fall under activities in the ICF scheme.

DR. HURWITZ-GERBER: But specifically from the patient perspective?

DR. MILLER: Yes. Now, I think that that's fair to say as CMS. However, in addition to that, what we are trying to bring out is the fact, is the questions about caregiver burden. We are quite

00129

aware of the questions of caregiver burden, so whereas it's not necessarily a coverage consideration, we do believe that it is in the interest of our beneficiaries to certainly consider that question during the study if it is appropriate.

DR. HURWITZ-GERBER: Thank you.

DR. DANIS: I wanted to particularly thank Dr. Duncan and Dr. Studenski for some very coherent presentations about the direction we need to go in and I wanted to ask both of you about your thoughts about measurements that will make it feasible to get large enough amounts of data to have any policy inferences. It seems like it's going to be very hard given how diverse the presentation of stroke patients are to get large sample sizes that you can get -- and it seems like you'd like to move towards data collection in the clinical setting where you actually observe a lot of patients and outcomes. And I'm wondering about what sort of measures you think could be useful in the context of clinical care that would be good and not too complicated to actually administer in that setting, and also ask you about, one other question, which is what you two think about duration of follow-up given the rate of progress in recuperation in stroke patients.

00130

DR. DUNCAN: In response to your first question about the clinical utility and feasibility of these measures, we have to go back to what the purpose of this panel is, and as I understand it is to look at technology. And if you are evaluating technology you should use the same standards that we might use in any FDA trial in which we come in with a conceptualization of what those instruments are and how we are going to apply them. And the range of measures that we've talked about are all feasible in the context of a randomized clinical trial to evaluate specific technology, and it's -- and again, I do a lot of consulting with drug companies so we're all, we can come up with a coherent battery that is not such a burden to the patient and can characterize the effect across the domains of the ICF model. So it can be done, it's easy, I've done it for years in a career, and it's not any more burden than any FDA trial. So that would be my response to your first question.

The second question, though, if you're asking it in terms of what is clinically useful and how you want to inform Medicare policy from a broader perspective, again, I would go back to the factor that we, to inform policy, we have access to a lot of

00131

information in Medicare files and records that give us very important outcomes, like fracture rates in this Medicare population, like rehospitalizations, and merging that with the number of days that they're living in the community can be done from a broader public health perspective.

And Medicare and CMS and AHRQ haven't even begun to tap that as it can be addressed in rehab.

There are other models that AHRQ and CMS are doing in cardiology in stenting, we're very involved in that at Duke, and we need to bring that same level of integrity and science into the rehab industry.

DR. KUBO: I would like to follow up on that question, because the issue is not what we can do in a randomized clinical trial which has CRAs, case report forms, databases, data clarification forms, and a thousand patients who are very well serviced. We're talking about collection of data in 10,000 patients where there isn't a CRA, a central repository of data or something like that. Is it possible to use these measures in a clinical sector rather than as part of an FDA trial?

DR. DUNCAN: Absolutely. Do you know how long it takes you to measure gait velocity and what equipment it takes you, and I can train a man off the

00132

street to do it.
DR. DANIS: So that's what we're asking.
DR. DUNCAN: Yeah. I mean, that type of index is that simple. If you think of the context of clinical practice in general rehab practice now, a patient is seen by OT, PT, speech and language, physicians and nurses. And if you do a survey of all those providers and you ask them how much time they spend assessing a patient, all of them will admit to about 45 minutes, 30 to 45 minutes. So in the course of seeing a patient in the multidisciplinary perspective, you may get eight hours of assessment with no consistent profile of that patient because each discipline brings in a different measure, doesn't use standardized assessments, and doesn't follow the patient prospectively with key indicators of outcome.
MS. FRIED: Actually, I had sort of a different question but in your presentation you made the comment, or maybe it was a slide that said can there be durable treatment without durable treatment or something like that.
DR. DUNCAN: Yeah, that was there, yeah.
MS. FRIED: So this is sort of a broader question because in my world representing

00133

beneficiaries you get your rehab therapy after an acute incident basically. The team comes in, they do their assessment, you get care for a certain number of days or weeks and then you're on a maintenance plan, and that maintenance plan depends on if there is a caregiver at home, it depends on so much. And so, this is probably beyond, although I don't really think it's beyond the mission, because you talked about challenges to research in the field of neurorehab and it seems like rehab goes much longer than that short period. Can you tell me if there is much research on, I don't want to call it maintenance plans, because maintenance plans in the Medicare world means you go on the plan and nobody helps you.

DR. DUNCAN: Yes, I understand that, and let me tell you what the challenges are. Of all the evidence that was reviewed, and I'll speak from physical recovery because I know a lot about it, none of the trials, constraint-induced movement, the walking recovery trial that I have going on in practice now is reimbursable or consistent with the Medicare reimbursement policy. So the level of evidence that's provided for intensity, frequency and duration under the conditions of the randomized clinical trial, as was mentioned by Dr. O'Dell and

00134

others of task specificity and (inaudible) in duration is not compatible with the current reimbursement policy. So what we had to do in terms of sustainability is to think about building more integrated models of care, and we all know that Medicare cannot afford to pay for every level of intervention that we might need, but again, drawing on my skill in walking recovery, it is paramount with my intervention that I get the patient to the level that they're mobile enough that they can sustain a level of activity and then be integrated into more community-based programs. So I don't -- I'm not standing here to say that Medicare should or could sustain the interventions forever, it's not exactly a statin pill, but we've got to get them to the level of physical functioning that they can sustain their well being.

DR. ALVIR: This is actually for Dr. Studenski, and I think you were practically begging for this question. We all know about the treatment and ceiling effects for a lot of these outcome measures, and we also know about all the heterogeneity in this study population, and we also know that a two-point increase or decrease in a scale

00135

really means, or may mean something very different depending on where the scale is. So again the question which I think what you wanted asked was, are there, has there been a lot or enough item response theory or Rash modeling done on these outcomes that we have been discussing? And again, this is not that popular, because even the FDA draft guidelines and patient reported outcomes don't even touch these things, so could you enlighten us on that, please?

DR. STUDENSKI: You're probably aware that there's a large contract that's been let and I think David Sullivan is in charge of it. But again, the theory is that we have dealt with a paper and pencil world where everybody has to get asked the same questions, and particularly in an area like physical function, there is a natural ordered ness to difficulty that can be used to range find using more computer-based systems. So you ask a person if they can walk; if they can't walk, there's no point to ask if they can walk a mile or two miles or so on, they don't walk, so then you want to know about how are their transfers, whatever. And if they do walk you may want to start finding out more. And so the idea is these tree concepts and these are implemented using computer logic sequences.

00136

And there, as you know, is a large major national effort to pool items from endless sources and come up with essentially as I understand it, the new generation SF-36. And I just think that because one of the major areas that's being developed is physical function, that that should, and you know, that should be integrated with where you're going, so I think the old days of the Barthel or any single item, an instrument like that, are about to be over, and you will be able to check a further range and you will be able to do it much more quickly. Another person who is doing work in that area that you know well is Alan Jette, who has item banks and is publishing in that area.

DR. ONDRA: I have a question for really anyone, but perhaps the tech assessment people. It seems to me that as I was reading through your assessment, the real problems that we're having are really fundamental. We don't have ideas in terms of what is baseline treatment to compare. If you're doing an RCT it's a little bit easier, but in observational studies you need a baseline to compare to to add a specific treatment, and what is that baseline? And it also seems, am I correct, that there's not a lot of disease-specific outcome

00137

measures, is that correct?
DR. OREMUS: Well, there were two parts to your question, so it seems that in some areas certainly there is a bit of a deficiency as far as the methodology goes. This goes to certain evaluation criteria where we're more deficient in terms of their methodological strengths than other evaluation criteria. But having said that, some studies were also very strong and some studies also were not very strong. So there really is a lot of variance in terms of methodology.
As far as specific measures, that was certainly one area that seemed to be lacking from a methodological perspective, is that many of the studies went and took off-the-shelf measurement instruments and used those in their evaluations, precisely because there wasn't any firm guidance as to what they should or shouldn't be using. So often that's what they did is they took something generic, and it's really inappropriate to use a scale just because everybody else uses it. What really has to be done is you have to assess what do you want to measure, is an existing instrument appropriate enough to measure what it is you want to measure, and does it have strong psychometric properties in your

00138

population.
DR. ONDRA: And without that you can't really calculate an MCID?
DR. OREMUS: Well, the minimum clinically important difference can certainly be calculated, it's the meaning behind the difference that is very important. And certainly if you're using an instrument that is not psychometrically appropriate in the stroke population, then what you calculate is not going to be a valid measure.
DR. SATYA-MURTI: You mean to say there is an MMCID, meaning behind minimum clinical.
(Laughter.)
DR. OREMUS: Yes, there are different ways to define what is, philosophically speaking, what is a minimum clinically important difference. But once you have your definition, your understanding of what it should be or what you think it should be, then it's certainly important. For example, in my opinion it's the smallest important difference that you would want to see that is clinically significant. It may not necessarily be the difference that everybody would consider important, but from your perspective what is the most important clinically significant difference.

00139

And that is obviously going to be dependent on the scale. If you're looking at a scale that measures change based on a point score, what is the minimum number of point change on the scale that's important, and that is in a sense where the difficulty lies. Is a two-point change clinically significant, often we can't answer that question because we don't know what a two-point change means clinically. So if we don't know what it means clinically, we can't understand if it's the minimum clinically important change.

DR. SATYA-MURTI: Thank you. Dr. Pauker had a question.

DR. PAUKER: This question has four parts. It's not clear to me why the issue of stroke is different than any other chronic disease that has long-term and short-term issues, and what I mean, do you mean that we need to think about this special and why don't we look at it with other chronic diseases to make it fair.

Secondly, it wasn't clear to me as to whether we're talking about effectiveness or comparative effectiveness. There is one slide that said the best available to use and there was a slide that says it could be varied to a placebo to define

00140

what kind of, we are picking the comparator, what kind of comparator we're picking.
Third, there's clearly lots of things that can be a very beneficial placebo effect, so if you have a study design that's A compared to A plus B, which is one mentioned to you a lot, you have to add a placebo to A, you need to compare A plus something to A plus B to see the effect of B, and I didn't see that mentioned in there, and I would like comments about that.
Finally, it appears as I've listened to lots of these things that there is a very broad set of potential outcomes in patients with stroke, so it doesn't make sense to have a single scale that covers all patients, not just what their deficit is but how bad their deficit is. Is there some large scale, because most of these things seem to be relative to differences in severely impaired people whereas in terms of functionality of a minimal difference at the high end of the scale may not be picked up well, and that may be very very important for integration into society or a job or whatever else, and I didn't see any comment about that. Tell me do you expect to have a single measure across the board, or do we need to have different measures for different variations

00141

of stroke?
So those are four questions.
DR. OREMUS: I will try to address each of the points. Regarding the first point, I certainly think that there is a certain amount of transferability of the issues that we're talking about today to other chronic disease areas, but I really can't comment further on that since we were focused only on the stroke aspect of these methodological issues.
The second point had to do with comparator treatments, and certainly that is one of the most important issues when you're evaluating any sort of therapy, be it a stroke rehab therapy or any therapy, is the validity of the comparator. And definitely one of the issues that we addressed in our technology report was whether or not to include studies without a comparison group, and we felt it was necessary to only include studies with a comparison group, because we feel that in order to evaluate any technology, any stroke rehab technology, you need to evaluate it against something. So certainly it's important to have a comparator treatment and it's important to have a quote-unquote valid comparator treatment, something that may be the standard treatment that

00142

you're seeking to improve upon or some other treatment that is used in the population of interest. The third point was placebo effect, that certainly is an important issue to consider in any study, especially in stroke rehabilitation where you may have other things happening in the background, it's important to bring those things forward. So definitely placebo effect is something that researchers in the future should be considering when they are designing their study. It's a methodological issue that they need to build into their design and certainly it's an issue that should be addressed in their discussion if they feel that there may be some effect on the result. So it's definitely an issue that needs to be addressed.

And what was the last issue?

DR. PAUKER: The last one is the single method when they can't cover the broader scale of potential disability.

DR. OREMUS: That's right. Some of the other presenters today may be better able to address that question. I think right now we're at the stage where we realize that there is an issue with the current crop of instruments used to measure outcomes in stroke rehabilitation, and so the first step is to

00143

recognize the issue. And then the second step is to really address the points that you've raised about whether we can have a global measure or we may need certain individual measures for specific issues. And I think that now that we've recognized there are problems with what's being done, the very questions you raise are the next set of issues that we may have to address in this field, and some of the other presenters today might want to expand upon that.

DR. SATYA-MURTI: Dr. Foley first and then I will have Dr. Roth after that.

DR. FOLEY: I was just going to ask a follow-up. Can you actually have a placebo effect in this particular disease entity where any amount of stimulation or stimulation, whatever is potentially having a therapeutic benefit consistent with what I saw in some of the slides with Dr. O'Dell about trying to get durable effect with durable treatment, and actually any amount of stimulation trying to stimulate plasticity and brain reorganization.

DR. OREMUS: I can't answer your question personally because I'm not an expert in the area of rehab itself, I'm a methodologist, so I'm going to have to defer that.

DR. FOLEY: But Dr. Roth is so I'm sure he

00144

can.
DR. ROTH: I actually had a question for all or any of the panel members, but Pam, you can start it out. And that is that you and others talked about comorbidities and complications as actual outcomes, listing (inaudible), rehospitalization, we heard about pain, we've talked about even mortality. I'm just wondering your thoughts and other presenters' thoughts about complications as an outcome measure.
DR. DUNCAN: Stroke is a chronic condition and most of the individuals who present with a stroke have the metabolic syndrome of diabetes and heart disease. I can tell you, again I'm unblinded to groups, but in my current trial that I have going on, there are a lot of competing comorbidities and a lot of intercurring events. I think that if we step back from stroke and think about aging and chronic conditions, we seem to understand now that the evidence is very converging that maintaining a certain level of physical activity and function may be the best magic pill, and that we may be able to influence recurrence of cardiovascular disease, diabetic management, and so we need to move broader. And that is the advantage that you have in Medicare

00145

and Medicare data, is to be able to look at the trajectory of these intervening comorbidities, rehospitalizations and recurrent strokes. So that has not been tapped from the rehab industry and should be carefully followed and I just think, I just use falls as a concrete example.

While I'm here I want to make one follow-up to Dr. Pauker's comment and something that Dr. Studenski said. I actually have, as I said, having a career trying to get the community to endorse more systematic measurements and not being very successful, I've taken a step back to say why is that, you know, why are we not there? To me it's quite simple. But I think it goes back to this question of clinical interpretability, do you really understand what is the meaning of your measures. I think, I always use blood pressure as a perfect example, we understand the range of normal blood pressure, we understand the risks with changes in blood pressure. And I think the real challenge is that in some of the measures that we've endorsed, and it's a particular problem with the new Rosch analysis and the item banking, I think it is the right methodology to be able to get the scope of function, but a clinician will never understand a logent score

00146

if they don't understand a change in gait velocity. So whatever we do, we have to understand the clinical interpretability of what we're doing, and that is the real challenge in some of these other metrics.

DR. SATYA-MURTI: Apropos to Dr. Roth's question, one aspect that hadn't been touched upon among comorbidities is post-infarct seizures. Many of these patients, as neurologists they might be actively undergoing postictal state or partial seizures, and if we include them inadvertently in one group or the other without knowing this is going on, and euglycemia is the other factor, we might actually bias the outcome one way or the other. If someone is being measured on a certain day with one of the indices when they're in a postictal state, that's really going to weigh it way down. So among the comorbidities mentioned, this is one that I didn't find particularly brought out except for some rare studies, so I wanted to put in a pitch in case someone is thinking of devising future studies.

DR. O'DELL: I wonder if I might just take a moment to address your placebo issue, and perhaps a couple other issues as well. I'm absolutely convinced there's a significant placebo effect because so much of what we do in rehab depends on

00147

engagement and participation. In so many folks, particularly in studies in chronic stroke, there's such a hope, there's such a desire to find whatever the next best thing is, the next step, that simply by being involved in a study and having the hope that something new can happen very well may motivate a patient that may have subclinical depression or psychological issues to really do more than they had done before. So I guess it's not exactly a psychological effect, but the better engagement very well could lead to functional improvements in a group that isn't receiving active treatment.

DR. SATYA-MURTI: You mean like a Hawthorne effect?

DR. O'DELL: No, I think it's probably more than a Hawthorne effect. They are involved, they are being observed and -- yeah, I guess it is, because they would behave differently, and by behaving differently and perhaps being more engaged in the rehab therapies that are being provided, they would put themselves in a position to benefit more from that.

DR. ROTH: For some patients, just being around the therapists and the clinicians is very beneficial, even if they're not doing any of the

00148

technical skills that we're talking about here.
DR. ONDRA: This really goes to a question that I wanted to ask all three of you, and that is the issue of blinding the patient to the therapy. Please educate me because this isn't my field of expertise, but it would seem that you could do that. You can't blind the therapist, but the patient doesn't know what therapy they're supposed to be getting in standard treatment, so if you add in an additional, I would think that you could blind the patient and get rid of some of that placebo effect.
DR. DUNCAN: Absolutely. You have to have a comparator control because it is beyond the Hawthorne, it's this idea of social engagement. Not to (inaudible) to give specifics, the CIT trial which I used as a model for outcome measurement was not the model for how you should select a comparator group, because rehab itself does require that. However, you can select a comparative intervention that, which they have to get the same exposure and to be quite honest, that you have some placebos that might work as well, the patients don't know and oftentimes the therapists don't know. So we can select comparative interventions that may not be quite as task-specific and could be an effective control, so yes, they can

00149

be done. You cannot rest on placebo.
And the other issue is you cannot compare it to usual care. The variability in usual care in this country for stroke survivors is phenomenal, and the things that you have to be able to control exposure to in an intervention.
DR. SATYA-MURTI: Dr. Pauker first and then Dr. Gerber.
DR. PAUKER: I want to for a moment take the other side of placebo effects. Placebo effect is actually a good thing, it certainly helps a lot of patients, so you don't want to discount that as a bad thing. We want to engage patients, motivate them, and many of these therapies are specifically designed to increase patient engagement and motivation because that can have enormous placebo effects.
On the other side of that, the flip side of placebo is that, we call it nocebo where I come from, and it was developed to talk about the adverse effects that happen by expectations. And I expect that in clinical trials and studies that nocebo effect may also be active and happening. So we need to think about both the positive placebo side with a P, and the negative nocebo side with an N, both can be conceivable.

00150

DR. HURWITZ-GERBER: This question is for Dr. Studenski. It's a follow-up on what I heard you say this morning, very enlightening to me about opening up opportunities for patients to indicate their preferences but attaching some sort of value to it. In other words, oncologically speaking you've got chemotherapeutic opportunities and then you make a choice based on number of bed days versus fewer bed days, et cetera. And that from a meta-question approach started me thinking about rather than coming up with single measurement tools that we could agree upon, six-minute walk time, group strength, Fugl-Meyer, how would you approach selecting not which measures, but how would you approach selecting the proper panoply of measurement outcomes for a process as complex as stroke?

So we within the ICF, for example, we have a number of domains, we have a lot of choices of selections within those domains. Some of them are very proximal to what we think the pathophysiology is and some are very much about patient choice, i.e., participation and that. How would you make a menu, if you would, based on a model such as the ICF that might help us choose an appropriate selection of outcome measures?

00151

DR. STUDENSKI: I think that is a really interesting and challenging question. I think I try to myself remain humble about the measures that I like the most because they're probably driven what I think is important, and that might not be what any particular patient thinks is important. So I like gait speed a lot but, you know, I've had people say listen, I've got one of those scooters, I don't care. So one scenario might be to say we were speculating, you know, could you have this Chinese menu where you say there's, you know, based on the kind of aspects of stroke impairments that are present in this patient, here are a set of reasonable impairment level measures, here's some reasonable activity level measures, here's some reasonable participation level measures, and is part of the process if there are several, to engage the patient in a discussion about their, you know, which taps into what's important to them. So I think that might be an element.

DR. HURWITZ-GERBER: Does that take you to the issue of meaningfulness? That's kind of where I'm going with this question. Without getting logent, you know, Dr. Duncan was talking about getting a number at the end of all of this, or in the

00152

SF-36 which gives you a number, but it's awfully hard to use that as an outcome that either leads you to treatment, which might be one issue, or that shows you the effectiveness of your intervention. So I'm trying to see if something like that is getting you close to the meaning.

DR. STUDENSKI: Right. So you're incorporating patient values but still trying to stay based in something that has other than space in terms of measurement. You know, the challenge with the balance is that there are social values that you're trying to incorporate into your decisions that say I'm not going to make everybody happy, right? So we, I think Pam and I run into people who say well, you know, my mobility goal is I want to be able to go out and run again, and it's not going to happen. So that there does have to be a balance between what is a reasonable societal expectation of a treatment goal and a patient's, and that was why I was trying to think of a way to incorporate both. And I think in terms of estimates of the magnitude of change that, you know, trying to have a foundation of patient values but then come up with something that's relatively consistent so it can be applied is where I'm trying to find the balance.

00153

DR. SATYA-MURTI: Dr. Grant has been waiting.

DR. GRANT: My persistence has prevailed. I was struck, this is primarily for Dr. Studenski, but I was struck from the technology assessment that, the number of instruments that had minimal clinically important differences. And if I could just make a quick comment I think, you know, these different metrics, there's a minimal clinically improvement I sort of like, how much the patient improves. There's a more detectable difference, what statistic you can find. But there's also worsening too, and all those are different quantities and need to be distinguished.

But from a policy perspective and evidentiary perspective it seems to me that, not seems to me, I do strongly believe that that is the quantity, that that is the benchmark that one has to reach to be able to show a real benefit. Now it's going to vary among individuals obviously, because that quantifies evidence in a way that we can understand in terms of how much benefit has accrued.

So to my question, though, what have been the barriers here in terms of defining that, because in terms of gait speed you seemed to point to it

00154

directly and make good points about it. And for what measures do you think that it's feasible to define such a threshold and what are not, and where might be the role.

DR. STUDENSKI: I think that it's a field that's rapidly evolving and as we do the work, we discover more challenges. So for me, for example, the problem with the distribution-based approaches are that they, one, assume symmetry, they say they're based on standard deviations as if the curves were symmetric, so they think improvement and decline are the same. They are also sample-dependent, right, so if you're calculating distributions, it depends on the distribution and the sample. So there are weaknesses and strengths.

Anchor-based methods, one of the things that we're really struggling with right now is that there's two main ways to do anchor-based methods. One is you ask a person about their state now and you ask a person about their state later, my mobility is excellent, very good, good, poor, and people have improved or declined based on how they have changed that rating. The other is to ask a person if they've changed, so my mobility has improved, it declined. And we were shocked and dismayed to find out that

00155

when you ask both questions twice, they don't relate well to each other, which was incredibly depressing. And there's a phenomenon where people recalibrate. And again, I think Pam and I have seen this for a long time, which is you ask a person how their mobility is and they say it's really good, and then you put them in a fitness program and you come back and they go, now I know it was very good back then. So both times they're saying pretty good but they're also saying they're improving, and there's a scenario I can do about decline the same way, that experience alters your perception of where you were. So I think we're working a lot now on how to get to the next step, and these questions of are these magnitudes different depending on where you are with some of these measures. You know, the gait speed low and high, certainly from many of the self-report scales you can't assume that the gains are smooth across the scale. But again, I think that that kind of stuff, one of the upsides is you can do a lot with observational data or secondary analysis of clinical trials. Certainly if there are scales that are being used in multiple small studies, you can certainly learn about the relationships between these measures as you try to calibrate meaningful

00156

change. It wouldn't even matter what the intervention is, you're just trying to look at how people perceive change, and so from a research point of view there's probably a lot of opportunity to pool analyses of data on performance and self-report measures.

And clearly, there are effect modifiers that we're just starting to look at. So you know, depending on culture or mood or many other things, some of these things may vary as well. I think the thing that I find heartening, because that was a lot of challenges, is, the one I know best is gait speed, is just how much it keeps coming out the same. I mean, I'll start throwing all these problems at it, what if I do it this way, what if I do it that way, what if I do it with this sample, and I just keep coming up with that .1 meter. We're worried about the decline in improvement, it's coming out the same both ways. So some of these problems are answerable and I think there may be measures that are reasonably robust to a number of these concerns.

DR. SATYA-MURTI: All right. Thank you. There's two others following, and then as a reminder, we have about 15 minutes left in Q&A.

I'm very impressed about the need to

00157

incorporate caregiver other than professional caregivers, such as family and friends. So that being the case, I was wondering if there has been any attempt at crafting an index that includes in the universe of evaluation of patients, caregiver input. Have they given up and gone to part-time, a spouse or a son or a daughter, or have they had to completely change jobs or go to night shift. So as I noticed, there hasn't been any concerted attempt at that, although that ought to be part of the global evaluation, is it not?

DR. DUNCAN: Well, of course I do believe that we need to look at caregiver burden and there are major implications to the family, not only -- again, think of the Medicare population. The Medicare population isn't usually the group that goes back to work but what we see is that, and we've done this research actually, that shows that it affects the health of the caregiver. So the heavily burdened caregiver declines in health and becomes extremely depressed, so that's another cost to Medicare.

DR. SATYA-MURTI: We haven't got an index yet to give some numbers to this, we haven't attempted any quantification of this yet, have we?

DR. STUDENSKI: I think there's extensive

00158

literature that's actually more about cognitive impairment, and you'd have to tell me where in physical impairment it is. But you know, things like unpaid care hours. I mean, it's a very tangible issue that I think may be very relevant. I mean, it's not just spouses, it can be daughters, I think it's very quantitative, to be able to estimate informal care hours per day or per week, it's a simple metric. And they certainly, again, psychological and health burdens on caregivers as well, but I think some of the simplest would be just, not just measuring paid care but unpaid care.

DR. ROTH: There are several caregiver burden scales and this literature is emerging right now.

DR. SATYA-MURTI: Ms. Richner.

MS. RICHNER: Well, I'm the eternal pragmatist, and I'm trying to pull some of this together in a sense to understand. Dr. Gerber, your question was right on. The issue from my perspective, and I need clarification from CMS again, is that you are grappling with the issue of having decisions that need to be made on technology and drugs, for instance, sort of that acute care kind of decision-making about, and how it relates to

00159

neurorehab activities. And I think your arena within the coverage group is not about necessarily healthcare services and rehab services over time. And so this issue of clinically meaningful difference is very very important in that you have to have the measures that the scientific community and all these, you know, very bright people that come here in speech pathology and all the different multidisciplinary areas, can come up with the instruments that clearly can capture clinically meaningful differences, and then you can decide to make a decision about whether or not this is something a Medicare beneficiary will benefit from. So I'm having a hard time here with some of the theoretical discussion in that we're mixing up so many different parts of your needs. We need to make, help you to have a tool or have a variety or a plethora of tools, or I was excited about this toolbox thing at the NIH and how, is that going to be used. When we go to FDA, for instance, I hate to bring up that feud again, but that's where it starts, that's where we have to come up with a conclusion about what is a valid instrument to measure whether the thing is safe and works. So all of that comes together here and, you know, I just want to make sure

00160

that we're pointed all together in a way that you can use.

DR. JACQUES: But we have a lot of needs so any help we can get is always appreciated. The dilemma that we often face, and it's been a while since I've seen a large volume of stuff specifically about neurorehab or stroke, we've obviously had a lot of other topics that have taken a lot of interest in the last couple of years. So part of the reason for convening you all today and having this technology assessment is I have this sense that there is this looming wave that is going to be sort of washing on shore over the next few years, and I would like to be in a better place in terms of being prepared to deal with that than I think we might be without some informing both of us and of the community. And it's not uncommon for people to come into us and say well, you know, I have this new gizmo and I would like Medicare coverage for it. And we say well, what kind of evidence do you have? Well, I got three trials, one has 14 people, one has 20 people and one has 32 people, and they all used different outcome measures, but my P value is great, so why don't you cover it. And as we all I think clearly understand, that's an extraordinarily

00161

challenging piece of evidence to try to make confident conclusions about. So to the extent that those investigators frankly could have invested their time and energy and their research subjects' time and energy in doing it better, however we want to define better, I think the patients are better off, they are better off, and in fact we're better off if we're looking at better evidence rather than worse evidence.

To the extent that some of those hurdles may be very difficult to surmount, possibly for reasons that are maybe peculiar to the stroke population but maybe more generalizable, we'd like your advice on how do we kind of mitigate some of those shortcomings in the evidence. Because, you know, the bottom line is that if a Medicare beneficiary is going to be better off with something than, I'll say she because most are women, than she would have been without it, then it's in my interest to advocate for that particular technology. On the other hand, if we have something that there is no reasonable expectation that that beneficiary would be better, and in the meantime pursuing this wild goose chase for this beneficiary would deprive her of the opportunity to pursue something that might have a

00162

much better likelihood of helping her, then I think there's a harm there that we would like to avert. So I guess getting back to my introductory comment, I realize it may be very difficult or impossible to get our arms completely around this, but even if the results of this end up being, okay, people are now aware of, maybe they don't know how to solve the problem but at least now they know there is a problem and they need to try to address it in their protocol so that maybe instead of a glaring issue it's a, well, okay, it's not perfect, but we can still get around that.

DR. SATYA-MURTI: That's a good point. Dr. Kubo.

DR. KUBO: Is it permissible to ask Mr. Mullen and Mrs. Wagner a question?

DR. SATYA-MURTI: Sure.

DR. KUBO: You gave two very clear presentations and objections, but I think we all agree that measuring outcomes is very important. You were somewhat critical of CMS in proposing certain outcomes measures, but I didn't actually hear your alternatives, and I'm actually asking you for sort of leadership by example. Do you have an example where you've taken, measured an outcome in a population,

00163

not what CMS mandated you to measure as an outcome but what you chose as a reasonable outcome, and demonstrated to your satisfaction that this therapy was either good and that you now use for all your patients, or bad and that you no longer use.

MR. MULLEN: From the ASHA perspective we don't really control what patients receive and so we don't make the clinical decisions in that sense. We certainly are working on a series of clinical guidelines based on studies using a number of measures. One example of such a measure that we think is important would be the national outcomes measurement system for speech language pathology, which is actually a measure that has been endorsed by CMS in the past. So that would be one example of a measure that's widely used, a lot of psychometric work has gone into it, but it's nowhere on the list of measures that were cited here. But in terms of making decisions about treating patients based on that research, you know, at the association level we don't treat patients, so perhaps Ms. Wagner has a perspective.

MS. WAGNER: Actually I work for Erickson Retirement Communities and we have large continuous care communities around the nation, and we have been

00164

looking for tools to use to measure our outcomes and benchmark against national standards. Our speech pathologists are using the NOMS, the tool he just described, at all of our facilities, and it's an easy tool to use, it's been in use for ten years, they have ten years worth of data. And those individuals or companies that participate in NOMS, it's free if you are a speech language pathologist, and we are able to get benchmark data comparing ourselves to other speech pathologists across the nation and their outcomes. And it's helped our therapists, knowing which areas they need to focus their skill sets on a little more because they might not be quite at that benchmark level for a certain area.

As far as physical therapy and occupational therapy, as you know, there are, or as you may know, CMS recommended four basic outcome tools, one of which is the AM pack, and that is a tool that we're seriously looking at as a company to use for our needs as we go forward. But we are hesitant because there's a cost involved with that, we're hesitant to make that investment since we don't know if CMS is going to recommend certain tools to use as we go forward. And if we invest all of this money into a specific tool that's not going to be

00165

used, that will not be a very smart decision. However, it looks like -- nothing has been created like the NOMS or that has filled that gap. The FIM test that is listed on your list and referred to in one of your questions really doesn't have adequate information on it to measure what a speech language pathologist does.

DR. HURWITZ-GERBER: What is the NOMS?

MS. WAGNER: It's the national outcomes measurement system, NOMS, and that has been around for ten years, it was created by ASHA and a whole panel of people. Rob can explain it in greater detail than I, if you would like.

DR. SATYA-MURTI: Thank you. As I understand the custom and tradition here, maybe we can take ten more minutes for any remaining questions both to formal presenters and public commenters, and then we go into panel discussion among the panel members, so ten more minutes of any pressing questions.

DR. PAUKER: Could I go to the last public comments, which raised some questions about pain? Pain treatment is an extremely important piece. Did any of the measures put forth prior to that comment, did any of them include attributes of how much

00166

discomfort a patient is having and whether the patient's discomfort is in some way affecting the functional status?

DR. SATYA-MURTI: Pain measurement, does anybody want to take that?

DR. PAUKER: Yeah. Pain strikes me as an interesting piece and I didn't see the others mention it.

DR. STUDENSKI: I think Pam can probably speak to stroke-specific things, but you know, all the global quality of life measures, you know, SF-36, they all have a pain element in them, absolutely.

DR. DUNCAN: We usually use the McGill pain scale. Pain is not that common in stroke, it occurs under two major conditions. The most common one is shoulder-hand syndrome, which is very painful, and in that we always endorse a pain measure. And the other one is if you have a thalamic pain syndrome, which is pretty unbearable pain. So at that point those, I don't know that they have been specifically validated in stroke, but we commonly use them in clinical practice in the presence of pain.

DR. STUDENSKI: And certainly you can detect changes in physical performance measures with interventions on pain, you know. So if you're in the

00167

arthritis world, which I know better, you know, if you intervene on knee pain, you change physical function measures, so they do pick up changes in pain.

DR. SATYA-MURTI: Dr. Pilley, you had a remark about pain, did you want to add to that?

DR. PILLEY: I think another measurement of pain is not necessarily in the symptomatic presentation or self reporting assessment of pain, which they might have a little improvement in pain but not to change their utilization of pain medication. And I think another more objective way of measuring pain is are they using the same amount of narcotics or pain medication, as well as are they improving their functionality, because I think that's where there is some risk in that. People may have a decrease in their pain because they don't completely understand what a pain scale of one to ten is. I mean, you know, I do some occupational med stuff, and people come in and say I've got a pain of ten, which really means you're in bed and you're receiving morphine, so on and so forth. But they may rate it as a seven and then say well, I have a five today, but their utilization of narcotics may have disappeared completely. So that's a significant

00168

beneficial outcome.
DR. SATYA-MURTI: All right. Dr. Miller had a comment about EuroQol incorporating pain measurements.
DR. MILLER: Yes. On the EuroQol measurement which is purported, or which has been studied in stroke and is purported to be valid in those patients, there is a pain subsection.
DR. SATYA-MURTI: Okay. Dr. Danis had a question.
DR. DANIS: I wanted to ask Dr. Studenski, it seems to me we're heading in the direction of having these expanded scales that focus in and it seems like inevitably the best approach to go. Is it ready for prime time in terms of trying to understand what it means clinically?
DR. STUDENSKI: I don't think so.
MS. FRIED: Actually I had a similar question about, they're called I guess computerized technology, and is that just geared towards locomotion?
DR. STUDENSKI: No.
MS. FRIED: So it's much broader, so can you tell us where that is with all the different, I guess occupational therapy and speech therapy and

00169

other measurements?
DR. DUNCAN: Well, there are major initiatives everywhere with this, but as Dr. Studenski said, multiple dimensional assessments from NINDS and NIH. There's groups like Dr. Alan Jette's group, we're doing this in cognitive functional and cognitive performance. So all that means is, if you remember when you took the GRE, the GRW is now computerized adaptive testing so if you can multiply, you know, two times two equals four, you don't ask somebody if they can do two plus two. So you find the level at which they can perform and you go up or slightly down. And so at zeros, you end very quickly on the items and the constructs in which you can function, and it's sort of like taking a ruler. But you can actually get the overall statement of performance with just a few items rather than a comprehensive battery of items. Again, the challenge for us is to put that into clinical interpretability right now, and it's a whole industry emerging, especially as it relates to physical functioning and cognitive assessment.
DR. HURWITZ-GERBER: I would just like to mention the NIH roadmap and in concert with that something called www.promis.gov. It stands for

00170

patient-reported outcomes medical information system. It's only about patient-reported outcomes, whereas Alan's work is much broader, it's more objective measures as well. But the methodologies there are spelled out beautifully on the web site and there are tools that you can use to help you determine whether or not you might create your own little personally created outcomes measurement tool. And it is, the coordinating center is at Northwestern with David Cella, and it really is taking off now as a very very important technology.

MS. FRIED: I actually had one more question, sorry, and it has to do with question three which we vote on, which says what is the minimum period of time that interventions be followed in order to identify a durable treatment effect. And I find it a sort of confusing question, so maybe someone can enlighten me. Does the zero to six months mean zero from like the moment there was some acute episode, or from the moment that therapy stops, and are we comparing therapies that last 20 days that they get in a snip, or a longer period of time? So if someone can help me.

DR. STUDENSKI: From my perspective it's important to build the answer to that around the

00171

natural history of the condition, and so it's not a single right answer question, right? And I also think that changing trajectories might be an interesting thing to do. You know, durable outcomes is a set of questions, but time, I mean, if you can get out of rehab faster because you can walk faster or whatever, that's a good outcome, or finish your home health more quickly.

So to me the idea of changing the course of recovery is the answer and whether that is acute, short-term, some of that might be weeks, some of it might be durable, in which case you have to go to the plateau phase. So maybe rather than having it be fixed on time it should be fixed on the basis of what you think the natural history of the condition is. Does that make sense? So we know what the natural history is, where plateau is likely to occur depending on whether it's a severe or mild stroke.

DR. SATYA-MURTI: Dr. Ondra, I'll have you ask the valedictory question and then we'll move on to the panel discussion.

DR. ONDRA: Okay. So the valedictory question is both a question and comment. I've sat on several of these MedCACs, and each time you hear what the problems and challenges are in the field and what

00172

is unique, and every field of medicine and health care has unique aspects that make common comparison difficult. Having said that, in a perfect world we really need to establish sort of measure, it seems to me, to allow us to measure relative benefit and value to both individuals and to society. I thought Dr. Miller's presentation at the very beginning was a great tone to kind of give a guideway on how we can get to that commonality, and I think unless we do that, it will be very difficult to answer those value questions. How's that for a valedictory address?

DR. SATYA-MURTI: Good point too. Next is a panel discussion among panel members. Whatever you either wanted to say or not wanted to say, this is a good opportunity, and you've given us almost 45 minutes?

DR. JACQUES: We actually may have given you more time than you need, but you guys might just be more efficient than some prior panels, so you're under no obligation to take all of the time if you don't think you need it.

There has been some discussion among various people that some of the questions, in particular questions four through seven may be prone to sort of being interpreted possibly in a different

00173

manner than was intended. It's not our intention at this meeting to say that only the following, whether they are those or others, only the following measures are appropriate or acceptable in trials that Medicare will look at. And to the extent that some panelists have felt, you know, possibly rather than voting on the question as it is, they might want to alter those questions or amend those questions in some way, that discussion could also take place during this period of time if you wish.

DR. SATYA-MURTI: Good point. I will provide you a slight alteration on questions four through seven so that it doesn't appear that we're endorsing one testing measure more than any of the others, so that will exculpate us from any bias. Before we go on to a discussion, I heard about Chinese menu, and I like to use the salad bowl metaphor, so culinary metaphors are really very good, we've got a common denominator, we all have to eat. And using that, I think the task today, correct me, Louis and Susan, but the task today is not so much as, are these current measures that were chosen based on the frequency of utilization in the studies, so are these the measures that you want, or would you craft something, or you have no confidence

00174

in any of these measures. So the idea is if you come across a device or a modality treatment request like Louis said, and from my own Medicare medical director days, that is very commonly the need that triggers literature search and discussion.

So if you come across a request for coverage and it's based on studies and you submitted those, what would you like to see incorporated into it and what kind of methodologies would you like? Would these suffice or would you have no confidence in any of these, would you want to do something different? Have I put that in correctly, Susan and Louis?

DR. JACQUES: Yeah. And I think if I could just add one little nuance on it, one way of looking at the questions would be these are only, for example, do you think that there exists validated reliable measurements for these particular things that people like to measure and report on in trials, and following below is simply an example of some if you want to consider them that way. So I think that's sort of a slightly different way of saying what Saty just said. I mean frankly, if you don't think that valid measures exist, one could simply say you have no confidence that this could be done at all

00175

with these or others.
DR. GRANT: Could I just make a comment, or first, Susan, go ahead.
DR. MILLER: I just wanted to say that in choosing these particular measures, it was done somewhat so that you could pick points of A versus C, B versus D, to give the pros and the cons, the advantages, the merits, the demerits of each of these, and then perhaps consider the characteristics of a better or best measure, if you will.
DR. SATYA-MURTI: So in other words, we have the liberty to say we have no confidence in any of these measures, so that will then be a setting for you to request they come up with something that is more global and more encompassing.
DR. MILLER: Certainly I think you can say that, but I also think that it might be worth your consideration to look at the measures and see what is good and perhaps not so good in each of them. Because again, they all have their usefulness as well as their disadvantages in certain situations. Some of them are more global measures, some of them seek to, may be a back door way perhaps of caregiver burden, at least in the way I personally think about them. And that was the point of choosing them, just

00176

to give some examples to play off of each other.
DR. SATYA-MURTI: With that in mind, one last statement and then I will shut up. Questions four through seven, before the word "indicator" on line two, we would like to modify it by saying, how confident are you that these outcome measures or comparable measures which have been validated as responsive, reliable and valid, and then go on to indicators. So in other words, that change would then indicate that not only these given measures or comparable validated measures would provide you confidence, and then consider them, all of them in not individually but as a group, going to Fugl-Meyer and so on.
DR. GRANT: Just a comment and sort of my picture of this forest here, because I think there are, it really is a bit of a forest. I think that there are a couple of issues here. One is in general for outcome measures short of death, most outcome measures aren't perfect, some are more imperfect than others. So the degree of uncertainty accompanying the use of one versus another will vary and will vary according to how it's administered, what the patient population is, how appropriate it is, and just a whole host of factors. So is there any one right

00177

answer and is there any one right salad bowl, is there any one right menu of items? I don't think so and I think it's probably a little bit, I think we're probably fooling ourselves to think that there might be.

So we're left dealing with uncertainties is the one issue and the, appropriately, you know, we want measures that have appropriate psychometric properties. You don't want to use something that's just random obviously, but none of these are. That part said, I think that, just to emphasize my point before, for the purposes of decision-making, informing at a policy level or even an individual patient, it is critical to have information conveyed, evidence conveyed in a way that's informative, that people can understand, they can intuit, although I'm not so bad with logents quite frankly, but you know, it's a scale, to make sure that the scales are integral.

And I think that that's where the major shortcoming is here, is that there are not well-defined minimum clinically important improvements. Now that may, maybe there's no magical numbers, but certainly we could say there's 20 percent, 10 percent, you know, outcomes reported in

00178

that fashion, so that we could make a statement, or CMS could make the statement, which people are in the business of doing, to say that we have this degree of certainty that this number of patients are going to benefit to this extent and we're probably going to be correct this amount of the time.

So to me that's how these measures, where their usefulness lies, and that's different from the other place where we had imperfect measures and some are more imperfect than others. So that, you know, as I said, unless we're looking at something like mortality, but we're not looking at mortality here.

One last comment too, but this just came to me, using adverse events here I think is entirely appropriate. I mean if it's just event-free, whatever time, or event-free, you know, the lack of answer is just a good a measure of accuracy I think, or effectiveness depending on where the study is being conducted, as manifest, absolutely.

DR. SATYA-MURTI: Dr. Kubo.

DR. KUBO: I have two objections to questions four through seven and I'd like to go over them in sequence.

The first one is, you can think of test A as being perfectly appropriate and effective in a

00179

certain patient population at a certain time with a certain intervention and much better than B. Conversely, I could think of a separate patient population and a separate intervention where B is better than A, and so voting just once is really not going to be helpful in that situation.

DR. SATYA-MURTI: That's one reason why we could consider them as a group representing motor and function abilities. I'm not necessarily defending it because I had a similar thought. And then number five talks about language, six about swallowing, seven about quality of life. So these principal domains among the ICF recommended compartments, so either one of these or something like that -- let's say tomorrow our workers come up, is that the walking test, let's say they come up with something equivalent or better. Would that be included among one of these measurements you would like to recommend? So I think we might consider them as a generic group.

DR. KUBO: Okay. My second objection is that the diversity of this panel is very useful in having a broad discussion about many of the questions and I think the discussion has been very rich and the different perspectives have been very helpful. But

00180

questions four through seven are really content-rich and really require expert opinion. As a cardiologist I deal with many of these different measures but not these specific ones, and I know the subtleties and so forth would be lost on me, and so I fear my vote would be inappropriate or potentially misleading.

DR. SATYA-MURTI: This is a good point. I think this was discussed several times, Susan may not be here, and that is why I think maybe we could consider -- I agree with you. I am familiar with some of these measures as a neurologist, but nonetheless I do agree. We don't know these subtleties as well as some of the primary workers in the area. But motor is a major part, cognition is another major part. So if these were broad divisions into those clumps and we could then go on to say yes, I don't want a measurement that only depends on quality of life and caregiver perspective but we do need something from the motor area, be it one of these or something like that. Is that the idea?

DR. JACQUES: That's certainly a possibility. It's certainly up to the committee, the committee would vote on whatever change you want to make to the questions, and it's not unprecedented for committees on occasion to say we just don't think we

00181

can answer this one question so we're going to skip over it. It's also not unprecedented to say we think this would work better for us as a discussion question rather than a voting question because we think this isn't a nail, and the voting hammer just doesn't work on this particular problem. And in fact the discussion that has ensued for the last ten minutes about this point I think has been very informative and if the committee feels that in lieu of voting confidence on questions four through seven, that frankly you just want to sit down and have a chat about them, it's fine with me, as long as you all vote that that's what you want to do with it.

DR. PAUKER: To continue that line of thought, as I read these before and I heard the presentations, I read through this material, I still don't know a lot about the measurement itself. So for all of those I would vote not true and that may not be what you want. I am stuck by one of the lines in the discussion, in one of the presentations, I can't tell you, it was Mark Pilley's presentation, about having no single outcome measure to capture the overall dimensions. I was (inaudible) but if I get to that, my vote in that sector, my colleagues' and other cardiologists' vote might be a random number.

00182

So having a measure of functioning is important, and that's fine, but picking up the particular measures, if there were an abstaining card, I would abstain.
DR. SATYA-MURTI: We could still give a number for that, Dr. Pauker, maybe we could give it three, which means that we're not sure, yes, they could be useful, or they could not.
DR. PAUKER: But by giving it a three says it is more than a one, and you know, giving a three as not sure is different than saying I don't know the difference.
DR. ONDRA: As a neurosurgeon, I completely agree with my cardiology colleagues here and couldn't have worded it better than the two of them, so I will just leave that go.
DR. JACQUES: Well, we achieved that at least.
(Laughter.)
DR. SATYA-MURTI: We'll just finish and then come right back.
MS. FRIED: I have sort of a basic question, I guess of CMS. Is CMS sort of adopting the ICF construct, which is fine, I just want to -- because I know at least one Medicare contractor had some discussion in one of the local coverage

00183

policies.
DR. JACQUES: I know there are some local Medicare contractors who have particularly adopted that construct. We're not suggesting that one has to use that construct and we're certainly not mandating it. We're simply presenting it as here's an example of one way of doing it. If the panel thinks that in fact -- and some of the public speakers have said, you know, wouldn't it be interesting if you could sort of integrate everything related to this field into one particular set of metrics so you could follow the patient whether they're in the hospital, post acute, through their disability, through their employment, whatever, and we could all speak the same language. I suppose that there are pluses and minuses to that as there would be to anything else. So we mention that not to suggest that you need to put your stamp of approval on it, but simply here's one way of trying to integrate this mess of 45 different walking functions and 45 something else.
MS. FRIED: It just seems to me that would be something worthy of our consideration, the use of the ICF. My sense is from what I've read is that it's a really growing use of that construct.
DR. HURWITZ-GERBER: It seems to me that

00184

we have a number of challenges with respect to answering the questions. One of them that is critical is that with respect to the individual and the evaluation of the individual, we need both subjective and objective measures, there's no way around it. If we're moving into an area in which we want to know about the patient's values and how they see their health function, we need to ask them. So one critical component is something which is both patient-reported outcomes as well as objective measures. I'm not willing to say, I don't feel confident in saying which ones, but that is a theme that I think CMS needs to hear from me.

The second is how difficult it is for people to agree upon a standard battery when the complexities are so great and the domains are so varied that we've started talking about Chinese and salad bars and things like that, and they are an interesting metaphor for dealing with trans-domain research. This is definitely trans-domain research. We're looking at, whichever model you want, the NAGI model or the IOM model, we are looking at things that are fundamentally inherent in an individual, not a group but an individual, as well as the interface between that individual and his or her environment,

00185

defined very broadly not in terms of society only but in terms of the physical environment as well. So we need measurements that address each of those domains. And unfortunately there's no one single one out there unless we get to five years down the pike when PROMIS and CAT and all these wonderful technologies help us through the maze, and we may be able to get there. And the third thing is the one that was brought up around pain, and I talk about that in my own mind about symptoms, how important symptoms are. Fatigue is the killer of rehabilitation. Patients who are fatigued are unable to respond to care and that is a very important variable in the mix and it's one that CMS in my view ought to be galloping along to try to figure out how to measure. It's difficult. Our cardiologists define it one way, our neurosurgeons define it a different way, our psychiatrists define it a third way, and on and on and on. So I do think that it is an area that has to be attended to. So we're now talking about the individual, the individual with respect to his or her environment, plus the symptom complex which has to be constructed in order to understand what the contributors are, both environmentally and within the

00186

person him or herself.
And another huge issue, which is how do you choose the comparators, what is this evidence? I would like to see our TA group, to be honest with you, take on that question, okay? Let's go after what is best evidence, how do you define it, what's the methodology, how are you going to recognize it when you see it, et cetera, et cetera. We can't in my view see each of these things independent of each other, they're all of a mix. And I don't think that helps you, unfortunately, pick one over another, but we know that there are some very good instruments out there. And it may well be that the walking speed or the stature per unit of time is in fact a surrogate for 15 other questions, but we just don't know that yet. So I think we're going to have to be somewhat flexible about this and not be so determined to pick one or two or three metrics that don't give us what we think we need in order to be intelligent about how to make very important, very individualistic, and often life-threatening decisions.
DR. JACQUES: I think you were more helpful than you thought you might have been with that particular response, because one of the things that is helpful for us actually is not so much that

00187

we would say you have to use this measure and that measure, but simply whatever measures you choose, don't only describe that well in the section of your publication, even if the editor is going to beat on you for page charges and everything else, put a reasonable justification in there of why what you picked is appropriate in this population at this time for this intervention, and make the same discussion with us when you come in rather than saying we used this one, this one and this one, so therefore we're wonderful.

DR. HURWITZ-GERBER: Leslie and Karen and I were sitting at lunch and we said wouldn't it be great if CMS met with the Archives of Physical Medicine and Rehab, the Blue Journal, the American Journal of PM&R, and said look, we need you guys, the editors and the publishers and the scientists to sit down and at least address the issue of what stuff needs to be done so that what's published in your first tier journals in fact cover these critical bases. That would be a start in the right direction as well.

MS. RICHNER: At least ones on the side of having to do the studies that they want, okay? What that means is that there is a responsibility again by

00188

the scientific community to come up with the logical tools that we can all be guided by, because what's going to happen is that it will seem capricious in a sense. CMS says you have to have this study and we want this, you know, gold standard comparator and this instrument, right? And then you go okay, well then, you go out to the community and the aphasia society and somebody else says no, that's not the one, no, that's not the one, so what do you do? So the challenge here is that we all want to come up with the same answer, and it's just extraordinarily frustrating until they can come up with some granularity, and that we can all work on getting to the same level. So to get it down to a pragmatic level, again, is fine with Archives and everybody else, but when is that going to happen, then we don't get there.

DR. DANIS: It strikes me that we are in a good position to try and say something about influencing the quality of the information base that will lead the policy decisions in the future and to the extent that the information is going to be useful for policy purposes, it has to be some kind of creation of a direction that creates some justifiable and valid uniformity to the data, because you need to

00189

be able to move in the future into thinking about what are interventions, clinical interventions that are worth paying coverage for. And it seems to me that in moving in that direction what we have generally said when you do quality of life literature and outcome literature is you have a broad array of different kinds of diseases that you tend to want to use standardized quality measures for. You need to have some measures that have some broad uniformity and some very specific measures for the particular disease you're studying. It seems to me we need to say something about that, that there needs to be in all these kinds of studies some highly validated and responsive tools that are used in all studies, and then some very specific measures for the given particular pathophysiology you're studying.

DR. SATYA-MURTI: Dr. Weiner, or Dr. Sloan.

DR. SLOAN: You know, echoing what you said and what Dr. Gerber just mentioned, I think one thing that should be considered is perhaps putting all these parties together in a workshop. You need to know what the patient values are, what the metrics are, how these things or what the standard for

00190

publication will be. And this has been something that had been approached in the cancer field with not perfect success but some degree of success, and at least we've been able to come to an agreement on what the areas of disagreement are and what things are accepted and where we need to go to make the next advances.

And so that would be the patients, you know, the people who are doing the research, the economists who have to figure out the metrics and the costs of these things, and that could be done, you know, within a two-day workshop. I suspect you're going to get a lot more out of that than having a panel that while broad in scope, you know, has a lot of members who really don't understand the subtleties of all these measurements that we're trying to make decisions about.

DR. GRANT: I was going to say some of what has been said before, but just to reinforce it, I think that some uniformity is absolutely a necessity ultimately on a pragmatic basis because eventually somebody comes with a device and says my device is better than your device, and if everybody is using different metrics it's absolutely impossible, and it causes conniptions for those of us

00191

who are trying to synthesize evidence from multiple sources with multiple outcomes. It makes it very very difficult to inform decision-makers, so that I think that at least some degree of uniformity is just, is absolutely essential.

DR. SATYA-MURTI: We do have to come back to four through seven in view of what we have all been listening to. Would the panel consider only for four through seven, we do know, we do not know, because I think those questions in some form need to be there because they address different domains. So is it acceptable instead of giving a quantitative grade, or even lumping them generically, but simply say we do know that these would be useful or we do not know, since the neutral question was also, as Dr. Pauker said, I'm not sure is not the same as do not know. That's too subtle, but maybe I don't understand the difference clearly, but we could say that we do know or we do not know? Is that acceptable?

DR. KUBO: Could I just say, there are some people who do know these instruments very well and have personal experience and understand the literature and the vagaries, so you could leave the question as is, but leave an option six perhaps,

00192

unknown, not sure, I slept at a Holiday Inn and don't really know anymore.
(Laughter.)
DR. SATYA-MURTI: Yes, I would prefer that too. Is the rest of the panel okay with that? So we'll add a sixth category in addition, this is an expanded language scale, six, we don't know.
DR. WEINER: Although we were on to something, that we feel like we know a good measure when we see it, when we have the right pieces. Now granted, a lot of those things are generic, the psychometrics that we've already discussed, but some we have all been educated today and perhaps we can comment maybe in discussion mode, and perhaps half of it has already been said, but I think some of your comments are on target, as were yours, Naomi. So I'm not comfortable with just saying yes, we know or don't know, but there are dimensions I'd like to see on the record perhaps, but I don't think we can vote on that.
DR. ONDRA: And I just had a comment for Dr. Gerber, and actually Dr. Sloan too when they're talking about a two-day forum. I think to do this would probably take more than two days, maybe several years of two days, but I think there is a format

00193

1 although I'm not sure that CMS is the right
2 organization to pull together a physiatry group for a
3 think tank, but probably the national organizations
4 are, and I would start there with answering these
5 questions. The CMS might be able to advise but I'm
6 not sure it's the right organization to --
7 DR. KUBO: They could bring them to the
8 table, that's for sure.
9 DR. SLOAN: CMS might be perhaps an
10 appropriate sponsor.
11 DR. HURWITZ-GERBER: I think CMS could
12 invite them to weigh in on this, and wouldn't it be
13 wonderful, could you imagine if the VA and CMS and
14 AHRQ and all of the regulatory agencies got together
15 and said yeah, we understand in concept and maybe we
16 can have some commonality, perhaps not the battery of
17 tests, but some common language so we could get it
18 out there?
19 DR. ONDRA: We desperately need some
20 national funding agency to look at clinical studies
21 and to fund that, and NIH, that is not their mission,
22 it's like three percent of their budget. So we would
23 either need a new institute within NIH or some other
24 organ of government to try to figure it out, because
25 it's hard to get a roadmap for the future when you

00194

don't know where you are now.
DR. JACQUES: AHRQ is here so we will have that conversation with them.
DR. KUBO: That suggestion for that interdisciplinary panel came from Mr. Mullen and I think that was a very good suggestion. We have actually done it just like Dr. Sloan, we've done it in cardiology many times. The key is to have all the stakeholders there and bring the FDA back into it as well, and it is a very useful exercise. It is one in which you do know fatigue because of everyone having their own opinions and being unwilling to bend to some of the others' opinions, but that discussion becomes very rich and I think gets you further down content-wise than where you will be today.
DR. WEINER: But one step further, future coverage, you must take that into consideration, so unless you're at the table it won't be covered, which is something I know that CMS has done in the future when it comes to outcomes.
DR. JACQUES: Yeah, we have amazing powers of attraction, it appears.
DR. SATYA-MURTI: Any other discussion on these issues? We are fast, aren't we? So, maybe we will move on to actual questions, I think Maria Ellis

00195

will tell us how to vote, is it the same procedure?
MS. FRIED: Before we begin, are we -- I earlier proposed we at least consider a vote, or have a confidence vote of the use of the ICF, is that something we could talk about, whether that's something --
DR. SATYA-MURTI: You have question 11 to put that in.
DR. KUBO: Are you saying sort of like, does this panel endorse the ICF concept of the three different domains as part of being important measures?
DR. FRIED: To be used by CMS as they determine coverage.
DR. SATYA-MURTI: Our task today is what are the gaps, so you might say a gap is the lack of consideration for any other alternative than the ICF, if that is your opinion.
DR. WEINER: The problem is that the tech assessment didn't really talk about it today, but I think we could endorse that as probably a good idea to talk about.
MS. FRIED: Okay, that's not a problem.
DR. WEINER: We might just make a statement saying we encourage CMS considering and

00196

investigating the use of the ICF as they review and move forward.

DR. JACQUES: And it's also important to remember that your comments to us are as informative as any votes.

DR. SATYA-MURTI: Are we ready to start? Do we need to flash the questions or we can just read.

Question one: There is the tendency to generalize stroke research to large heterogeneous populations. How confident are you that the strategies below represent meaningful comparators in observational studies?

A is protocol driven usual treatment versus protocol-driven usual treatment plus the specified intervention. So the specified intervention is the add-on here, and you want to start with confidence levels.

(Panelists voted and the votes were recorded by staff.)

DR. SATYA-MURTI: How about choice B, patients himself or herself before and after intervention.

DR. ALVIR: Could you clarify?

DR. SATYA-MURTI: This is the patient

00197

after the intervention?
DR. MILLER: It's not an N of 1. It will be looking at whatever he or she was doing and then you apply the intervention and take your observation, then you look at what the next patient is doing, you apply the same intervention, you take your observations.
DR. SATYA-MURTI: On the same patient.
DR. MILLER: No, on different patients.
DR. ALVIR: On the same patients.
DR. MILLER: Pre-post.
(Panelists voted and the votes were recorded by staff.)
DR. SATYA-MURTI: Okay. A slight variation here, patient himself or herself before and after treatment, then with treatment withdrawn and reinstituted as appropriate.
(Panelists voted and the votes were recorded by staff.)
DR. SATYA-MURTI: Where do we stand on non-protocol-driven usual care versus intervention?
(Panelists voted and the votes were recorded by staff.)
DR. SATYA-MURTI: Next is the non-voting question. It calls for some discussion.

00198

Large prospective randomized trials are uncommon in this field of medicine. Discuss how other study designs can or cannot adequately account for potential confounding factors such as: Natural clinical course of recovery. I think I'll take them all in order. Selection bias due to: Skill level of therapist; comorbidities affecting both the stroke etiology and course of recovery; ancillary therapeutic resources, virtual home/community environment; severity of illness. Differing assessment tools used across care settings, inpatient rehab, skilled nursing facilities, home health agencies, outpatient centers. Pre-morbid and cultural characteristics. Discharge settings and social support.

Since I'm not voting, I'll go down the table on this.

DR. GRANT: When you say randomized trials, are you specifically referring to specific trials where the patient is randomized, or there may be other randomization points such as sites.

DR. SATYA-MURTI: I think this is just the patients.

DR. GRANT: So we're only discussing patient randomization. That being the basis as a

00199

preamble, I think part of the answer to these questions is what we already discussed. I feel that the natural clinical history of stroke is so variable I feel very diffident about doing any of these trials. As far as comorbidities, I have seen stroke patients where the brain is relatively intact and there will be an infarct of myocardium or even intestinal bowel removal, so it's not a critical link (inaudible) so that being the case, these comorbidities have a great effect on the recovery itself which masses the comorbid difficulties, and so I feel --

(Inaudible portion due to tape failure.)

DR. SATYA-MURTI: Would you include a subset analysis?

DR. DANIS: I think you can have, when you have that (inaudible) complete diagnoses, sociodemographic data, characterization of the care setting and the interventions, and I think we would be way ahead of the game on everything.

DR. SATYA-MURTI: Thank you. Others.

DR. ONDRA: Well, I agree with that and there is some concern about collection being hugely expensive, and so this was really an EMR issue so I

00200

think that's a very plausible situation. What you don't want to do is look at and say we can't afford to give it because the data collection requirement is worth doing.

DR. SATYA-MURTI: And this collection should be initiated prospective to any subset analysis?

DR. ONDRA: Right.

DR. PAUKER: The question really comes down to variability, but in order to look at the effect of all these other things, each of these things has to be measured, and if you can't measure the variety of illness, there's no way you can make these adjustments, so one of the steps is that has to be developed as a measurable documentable outcome. I think that's very important. Subset analyses are important in RCTs so even designing them size-wise wouldn't be effective if we didn't have the right subject patients because of the subsets.

DR. SLOAN: You know, I hate to be divisive, but we're saying there are almost 800,000 patients a year that fall into this category, so why can't we put together a prospective randomized trial. Perhaps it won't be all double blinded because there are certain hurdles there and complications, but to

00201

spend huge amounts of money to collect data from sort of random and highly variable sources, from institution to institution, it may be very hard to really make any sense of that, so I don't know in the end what you're going to get out of it.

DR. GRANT: (Inaudible) registries, and I'm not familiar with the field as many others here aren't, but I think that conceivably I could see a useful place for a well developed, well conducted registry that collects the correct data to do an analysis of the natural history of disease under usual care, which is critically important to understand as a platform from which to perform the appropriately designed randomized controlled trial.

MS. FRIED: There is a huge warehouse of data that CMS has, is it ten years old, because we just got a home health initiative and there is just, I actually don't know, but my sense is like a treasure trove of data on these chronic conditions.

DR. ONDRA: Maybe, and maybe only because, what is the data integrity, so part of the question is the data integrity, and so right now what I said earlier is we need a better --

MS. FRIED: I thought it was claims data, but I may be wrong.

00202

DR. WEINER: Of course we should try to do RCTs when possible, and often they are not possible in a small group, and that's not the problem here, there are other factors that we won't go into. Secondly, if we're relying on nonrandomized studies, I think these capture (inaudible) one, but I think the same workshop that we were talking about could address the study design and covariants, and I think you're on the right track here.

And thirdly, you know, the learning organization as we move forward, I think there is a lot of variability and I would say that perhaps another workshop, I think that clearly there needs to be protocols in the EMR context and so that too, and they certainly would capture certainly the clinical aspects here if not the organizational.

DR. SATYA-MURTI: Dr. Roth.

DR. ROTH: The question asked about how we can account for those confounders and this speaks to the theoretical use of single subject design.

DR. SATYA-MURTI: You're suggesting that as a potential alternative?

DR. ROTH: Right, as an alternative to help account for these multiple factors or N of one

00203

studies.
DR. SATYA-MURTI: Would that have a measurable effect when you stopped therapy when you're doing nothing and then measuring a leftover effect from treatment period when you're going to non-treatment period?
DR. ROTH: Sure, that's certainly a theoretical possibility, but if it's a well enough designed single subject study, then there are ways to account for that.
DR. SATYA-MURTI: All right. May we move on to three? Maybe a show of hands might be the correct way of voting on this, what's the minimum period of time that interventions be followed in order to identify a durable treatment effect? Who votes for the --
DR. GRANT: May I ask a question out of order? I mean, I just had a little difficulty here and I think Dr. Studenski commented about these different natural histories, so for some of them zero to six months might be appropriate, for others it might be longer. It's hard to put anything into a specific category, that's just my take.
DR. ONDRA: I interpret this as from what we understand, correct me if I'm wrong, that any

00204

follow-on time is really looking at durability of the effect.
DR. SATYA-MURTI: Accounting for all that, if you were measuring it, would you stop at zero to six months? We need to answer that in spite of those shortcomings.
DR. PAUKER: What do we mean by durable, do we mean after therapy stops or how long the therapy continues? It isn't clear from this question what do we mean by durable effects.
DR. JACQUES: It could be looked at in two different ways. One would be, is the subject going to still be continued after a month that treatment effect is essentially seen and he was getting it. I think the way the question was initially conceived, though, was more if there was an intervention that has a beginning and an end, that once the treatment ends, how long should we follow the patient to see if the treatment had a durable effect, six months after, whatever.
DR. PAUKER: So which one?
DR. JACQUES: You could almost do it either way.
DR. PAUKER: You've got to pick one or the other.

00205

DR. JACQUES: I would have to say it would be how long after the interventional treatment period has stopped.
DR. SATYA-MURTI: So after cessation of the intervention, would the effect spilling over if you measured it for zero to six months and so on.
DR. DANIS: So this is a duration of time after the cessation of treatment.
MS. RICHNER: So your coverage decision would be based on if it stopped.
(Inaudible colloquy.)
DR. JACQUES: I mean, imagine it this way. Imagine there is some therapeutic intervention, exercise, or whether it's stimulation or whether it's something else that is not something that is clearly designed to last for a month and that, if you use a wheelchair and you stop using it, you can't get around anymore, there are -- the wheelchair doesn't work. Let's say you had some, for example a series of exercises or something else for let's say a month, it was designed to last for a month, that is in fact the intervention that is being marketed, do X for a month.
DR. ONDRA: It was an --
DR. JACQUES: If it was designed for

00206

chronic use, you're not taking it away from them.
(Inaudible colloquy.)
DR. ROTH: Well, you know, I think there are some treatments that you could see an immediate effect and then it wears off. Traditional exercise is an example where if you are not exercising often, they don't have a persistent effect, so it speaks to the idea that you would want to have as long an effect as possible, and some of these technologies or techniques are making a claim that they will show that there's, you know, the more durability the better.
DR. WEINER: Which of these numbers do you like? There are things you might ask my --
DR. ROTH: Again, it's a judgment call. I would say a year.
DR. ONDRA: There's differences, like for spine surgery --
(Inaudible colloquy.)
DR. SATYA-MURTI: After Dr. Gerber, you had a comment, and after that maybe we'll go to voting on this.
DR. HURWITZ-GERBER: I just want to say in support of what Elliott was saying, you know, if you've got a frozen shoulder and you're working with

00207

therapy, you're expecting that that shoulder's range of motion will come back and stay back, and you would imagine that a year or so would tell you how durable your response is. I think the concept is somewhat confounded when talking about durable medical equipment, this is not the same as equipment. Therapies, I mean maybe in terms of durable equipment concepts you've got a different understanding, but obviously given the confounders and given the nature of the process, it would be very difficult to come up with an opinion that was educated on my part about what is durable. I mean, sometimes it's the life expectancy of the individual, which may be two months. So I have to duck this one.

DR. SATYA-MURTI: That will be all right. I understand the reservations. Anyone else?

DR. KUBO: Part of it is depending on the intervention. For a surgical intervention, I would want that to last a year. Something that is less invasive, just requires one visit, if it lasts for a week or a month might be okay for me.

DR. ONDRA: But the costs may be the same.

DR. SATYA-MURTI: Okay. Does anyone think zero to six months would be sufficient?

Six to 12 months?

00208

12 to 18 months?
No one wants to wait for longer than 18,
all right.
Well, four through seven, that's the devil
in the details. What we wanted to do is modify the
question in the second line, indicator of -- allow me
to reread that. How confident are you that outcome
measures like the ones that follow or those that are
comparable, are reliable, valid and responsive indicators
clinical trials that aim to improve an individual's
functional capacity in the performance of ADLs/IADLs
and locomotion or transfer abilities?
We would consider all of those as a
potential comparable measure and vote on them as was
told this morning. Ready?
DR. HURWITZ-GERBER: If we feel they're
reliable and valid but not responsive, how do we
handle that?
DR. SATYA-MURTI: Give them a weighted
score in your own mind.
(Panelists voted and the votes were
recorded by staff.)
DR. SATYA-MURTI: All right. I think

00209

We're going to enter the same six choices for the next one. How confident are you that each of the outcome measures like those below or those that are comparable, are reliable, valid and responsive indicators of change in clinical trials that aim to improve an individual's functional capacity in the performance of language and communication skills? Aphasia Quotient of the Western Aphasia Battery and Porch Index of Communicative Ability.

(Panelists voted and the votes were recorded by staff.)

DR. SATYA-MURTI: Thank you. Six, how confident are you that outcome measures or comparable measures like the ones that follow or those that are comparable, are reliable, valid and responsive indicators of change clinical trials that aim to improve an individual's performance of swallowing? Coughing, choking frequency during a meal, videofluoroscopy, and we are taking them as a whole.

(Panelists voted and the votes were recorded by staff.)

DR. SATYA-MURTI: We're done with that. Seven. How confident are you that each of the following outcome measures are reliable,valid and responsive

00210

indicators of change in clinical trials to assess patient, proxy or caregiver perceptions of the patient's health and satisfaction with life and community re-integration? Barthel Index, Modified Ashworth Scale, EuroQol.

DR. HURWITZ-GERBER: And the proxy in here is not only a healthcare provider but could be a significant other.

DR. MILLER: Correct, it could be anyone chosen by the patients or anyone who knows the patient well, but it does not have to be a caregiver.

DR. DANIS: I find it a little hard to lump these, one is, Ashworth is very narrow and the EuroQol --

DR. MILLER: If I may interject here, one of the reasons for this particular question was to develop a discussion on whether or not these measures are appropriate as quality of life. They have, looking at the TA, they have been used as quality of life measures and what we wished to bring out here was whether or not they should be.

DR. PAUKER: I don't know if we should lump these, I don't know about these other ones, but if we could break those out, that might help.

00211

DR. MILLER: Yes, if that's your pleasure, certainly we can break them out.
DR. SATYA-MURTI: And we can use six for those with which you just don't know.
DR. DANIS: It just strikes me that our goal as you get more and more to the more subjective components, the capacity of a surrogate to reflect accurately on what the subject is perceiving is so much poorer, so it just seems to me, are we asking are surrogates good measures? I mean, it's a tough set of questions here.
DR. MILLER: I agree, and I don't want to put my views onto the panel, but that was also placed in this question for discussion. When we say a proxy, clearly what the patient perceives as his or her quality of life versus let's say what a family member perceives as the patient's quality of life may be two different things. Certainly there are studies of this question, perhaps most dramatically in the patient population of those with ALS in which, you know, watching their kids grow, et cetera, are for these patients, their definition of that which is a very good quality of life. In my teaching experiences, however, I have found that concept very difficult to be appreciated by my students.
DR. DANIS: Yeah. It makes me want to say

00212

that I would endorse EuroQol measures taken from subject, but be skeptical about EuroQol measures taken from surrogates. I mean, unless you're interested in hearing about the impact on the caregiver, and that some of these are not measures of that.

DR. MILLER: And that is some of the difficulties that have been raised in the studies of the psychometric measures, the individual testing measures.

DR. SATYA-MURTI: The proxy versus patient agreement was fairly good with EuroQol for motor indices but not for psychological well being, so it's got some merits to it. So we'll consider this again individually in view of the fact that Ashworth, I also think is more designed for spasticity, we use that for MS patients, and EDSS, so we'll take them individually. Do you feel about Barthel Index can be rated one through six, we're including six here, don't know? So we'll go with Barthel first.

(Panelists voted and the votes were recorded by staff.)

DR. SATYA-MURTI: Okay. Modified Ashworth Scale, is that a good indicator for quality?

DR. FOLEY: In general or related to

00213

spasticity?
DR. SATYA-MURTI: No, in general for the question, which would be a sensitive indicator of quality of life and community re-integration.
(Panelists voted and the votes were recorded by staff.)
DR. SATYA-MURTI: All right. EuroQol, what does the panel think about EuroQol?
(Panelists voted and the votes were recorded by staff.)
DR. SATYA-MURTI: Okay. Question eight, how important are caregiver burden and their narratives as indices of successful rehabilitation? I was thinking about it, and do we need to go to one through five, or just say very important, somewhat or not at all important?
DR. DANIS: May I ask a question about this?
DR. SATYA-MURTI: Yes.
DR. DANIS: It seems to me that whenever you're making these kinds of value judgments, it's whose perspective. I mean, are we asking when thinking ultimately down the road about reimbursement, do we want to be thinking about the broad family context, or are we saying the narrow

00214

question when we are assessing rehabilitation as an endeavor? Does it -- are we thinking from the patient's perspective? I think you could answer this question depending upon what perspective you take.

DR. MILLER: This question is meant to address it from the caregiver's perspective because it is their narrative. And it is, meant as a way to think of their narrative as a type of outcome measure, of the success or non-success of a rehabilitation method device, whatever.

MS. RICHNER: I just have a question. When I was reading the materials, the issue was about narratives versus a quantitative scale of some sort, and they dismissed the issue of narratives, that those were probably not used or they were too indecisive. So to me, I would love it if the question could be clarified whether or not you could use a scale for caregiver burden, because it obviously is extraordinarily important in stroke, however, it's to me, the problem is whether it's a narrative or a quantitative measure of some sort.

DR. SATYA-MURTI: That is a good question because when we were thinking of this we were not sure how validated and how, what kind of longitudinal experience we have had with these scales. We just

00215

heard that they are emerging and they are not fully fleshed out yet. Is that correct, Dr. Roth? Were you not the one who was saying --

DR. ROTH: There are many being used for --

(Inaudible colloquy.)

DR. MILLER: This question was meant to be a narrative index. The EuroQol was meant to be more of a scaled response of the patients and his or her proxy.

DR. WEINER: Does EuroQol ask about burden on caregivers, carers as they say? Then it's really proxy. So I would propose that we, how important would be reliable measurements of caregiver burden as indices, that's certainly what I would like to vote on.

DR. SATYA-MURTI: May we take it as narratives and validated indices separately?

DR. WEINER: We could use separately if you'd like.

DR. SATYA-MURTI: All right. Because what I'm thinking of is if they have emerged but not fully emerged yet, then are we missing out something by saying we would confine ourselves to known scales only? Is there an aspect of caregiver narrative that

00216

hasn't been captured unless we listen to it at large for a defined population?

DR. ALVIR: Well, they are being used now in Alzheimer's because it's very important there for the caregiver.

DR. MILLER: Right. There are also some in the congestive heart failure patients, for example, up and coming tech research. May I suggest that question 8.A be the narrative index and then B would be a currently validated scale.

DR. SATYA-MURTI: So first for narratives, do we have, are we going on one through five? If so, okay.

(Panelists voted and the votes were recorded by staff.)

DR. SATYA-MURTI: All right. B, of the same question, how about a more formal validated index?

(Panelists voted and the votes were recorded by staff.)

DR. SATYA-MURTI: All right. Question nine, how confident are you that these conclusions can be generalized to community practice settings outside the context of specialized treatment centers? We talked about efficacy versus

00217

effectiveness and these conclusions would refer to the voting that we chose to the foregoing points up to this point. So how confident are you that these can be generalized from special centers to community practice, one least confident, five most confident?

DR. DANIS: Are we talking about when a practice setting does a serious job of trying to do research, or are we asking how -- I'm not sure I understand. Are we just saying --

DR. SATYA-MURTI: Once an efficacy trial has been done and you have a publication and peer review, then we start using it in community hospitals and --

DR. DANIS: Oh, I see. And doing studies to measure effectiveness.

DR. MILLER: No, that you use the results of your study.

DR. DANIS: Oh, the results of the efficacy study in practice and how confident are we that we can translate, that we can assume that efficacy data is useful without having done an effectiveness study?

DR. SATYA-MURTI: Yes.

(Panelists voted and the votes were recorded by staff.)

00218

DR. SATYA-MURTI: I think many tech assessments do ask questions nine and ten, so it's applicable across the board. Ten is, how confident are you that these conclusions can be generalized to the population of Medicare beneficiaries, knowing that age and other comorbid conditions and so on?
(Panelists voted and the votes were recorded by staff.)
DR. SATYA-MURTI: And the last is a discussion item which we have done up to this point in various guises, but I'll read that and then we have some time, it's 3:09. What are the gaps in the current evidence on stroke rehab therapies in Medicare beneficiaries? We've already spoken about this but I'll start off with anyone who wants to lead. What gaps do you identify? And one was that ICF itself may not be the only game available.
MS. FRIED: I think what I said was actually sort of the opposite of what you just said, which was that I don't think CMS has really publicly used the ICF in doing an analysis or looking into their coverage decisions, and that's something that I know that some of the Medicare contractors have, but I don't think that CMS really has. And I was just saying that that's something that they really should

00219

think about given the prevalence of chronic conditions in the Medicare population.

DR. SATYA-MURTI: All right. Anyone else wants to fill the gaps? Yes.

DR. HURWITZ-GERBER: I think it's almost a repeat of what I've already said, but first of all I think apropos of what Leslie said, we need a model, and the model has to have the domains that CMS values or that your advisors help you value.

Second is, I think we need some exploration of the literature to help us understand what the criteria would be for selection of comparators in a variety of settings and explore what constitutes a reasonable way of selecting those.

In addition, we did talk about just recently in the last couple of votes of the fact that we have to have some data, or I believe we need to have some data on caregiver burden, what is that, which domains are included under caregiver burdens.

We've talked about potentially economics, we've talked about psychological, physical and health issues. I think we need to consider that as a very critical component.

And there were two issues that were brought up following the TA which are still sticking

00220

in my mind which I think need better clarification. For example, under one of the presentations the slide was entitled results, appraisal of quality domains. There were two areas of really significant methodological bias: one is baseline characteristics and the other one is cointerventions, and I think that that's a huge gap. I think we have to understand how to evaluate patients at baseline if we're going to measure incremental change, and in this field we absolutely have to measure incremental change because our RCTs are so expensive and so difficult to do, that that's a huge problem. And then the issue of deciding what are the cointerventions, are we talking about mainly pharmacological or non-pharmacological, or environmental. We have to begin to bring some systematic approach to how we're going to evaluate those quality domains.

DR. SATYA-MURTI: Yes.

DR. DANIS: I wasn't clear from the technology assessment as to what was the age range of the studies, but I think we need to think about the applicability, particularly as you think about the Medicare population and the fact that you are including people who are, you know, of vastly

00221

different ages. And as you get more towards the end with the frail elderly, what can we say and what inferences can we make about rehabilitation of patients with stroke as they become more and more frail, and as that fraction of the Medicare population expands, I think it's really important to know.

DR. SATYA-MURTI: Yes.

DR. FOLEY: I just have two issues. One, we're not a homogeneous population in the United States so that the burden of stroke and the issues related to various subcultures I think is very important and we have to look at it, and the burden on those types of caregivers. In Minnesota we have a large Mung population that has a lot of stroke issues, for example. They tend to be very close-knit. I think the patient population that I see that has a stroke and then the complications afterwards, they, the family and the family circle gives a lot of care to those folks to the point that they don't do anything for themselves. So that we might do some things in the hospital that gets them to a certain level and when they go home, they become couch potatoes, they do nothing physical, or stimulation drops off, so we need to understand those

00222

populations.
The second I would say is, and it relates to a lot of studies, there's a disconnect between those of us who are at the hospitals, community hospitals, who have a lot of patients who have this burden, and the academic centers where you have these study designers and many statisticians, and all kinds of folks who do a lot of studies but who don't have the patients. And unless we have a connection between community caregivers who have those patients and the academicians, I think the power on these studies are going to be lacking.
DR. ONDRA: I think Dr. Danis' answer was an important one, it affected how I answered number ten, because I interpreted that as to the Medicare beneficiaries as to a large group of them. I think your point is a good one, that Medicare beneficiaries is too heterogeneous a group to generalize to and you almost want to, to me, have that question to a large subset or the entire Medicare population, because the Medicare population of people over 80 is an entirely different group than the Medicare population between 65 and 75. So I think that it really depends on whether you interpret it to the entire Medicare population or to a large subgroup of the

00223

heterogeneous Medicare population.
DR. PAUKER: I'm going to take the opportunity to push my colleagues at Medicare where they probably don't want to be pushed. But nonetheless, I think it's important that we have good studies on comparative effectiveness that we don't have now, and I think it's terribly important that those studies look at costs. Now there are lots of therapies where the effectiveness is there but it's more, and if we don't begin to look at costs and if you don't begin to look at the cost of care, then more and more money is in Medicare which we can't afford, and we're gaining less and less. And I realize that this requires enormous political will, but that requires leadership and you've got to start somewhere, and this might be a place to do it.
DR. SATYA-MURTI: Dr. Roth.
DR. ROTH: A couple of research gaps. The first is the relative role of medical comorbidities as confounders of outcome and also the role of medical complications as potential measures of outcome cost.
And then the second is as we talked about, the ICF model, it's easy for us to think about the connection between body functions and activity level

00224

and participation, but that's not been very well studied, the relationships between them and what might mitigate against or support one predicting or being associated with the other.

DR. SATYA-MURTI: Yes.

DR. WEINER: I think building on what Dr. Pauker said, although I've heard eloquent arguments for the individualization and, you know, the person and the patient, this is by and large a societal population-based program of 32, 40, whatever it is, million Americans, and we must look at population-based measures. And moreover, everything is very atomized into fee for service or CPT codes or whatever. That's all well and good, but it's not really about study of one little CPT code at a time, but rather systems of care. I'm hearing, and I've learned a lot, a lot of very integrated care, systems of care, and usually as I've been on many panels, I've asked Kaiser Permanente or the VA or the UK, somebody that thinks about populations and doesn't worry about CPT codes, what they have done. We didn't hear much about that today, but I'm sure there are those in the stroke and rehab care community that do think about it in an integrated way, holistic way and societal way, and that's a lot more than research

00225

design, but I think it's worth mentioning and studies that capture some of those dimensions should be supported.

DR. SATYA-MURTI: Leslie.

MS. FRIED: I just have two quick comments. One I have to say at every MedCAC is that there is nothing in the statute, in the Medicare statute or law that says cost effectiveness, the law is medical necessity. And this has been a longstanding struggle with CMS, and at this point there's nothing about cost effectiveness in the law, so I just wanted to make that point.

The other gap I wanted to mention, actually I raised earlier, and it has to do with what is called in CMS terms as maintenance therapy. And although -- and it might be worth some study into, if somebody has an acute care episode and has whatever their initial therapy course, they then sort of get given a maintenance plan and sent on their way. And I think that there may be a gap, I don't know, there may be a gap in services like what happens with that maintenance plan because you have a maintenance plan but Medicare doesn't cover the maintenance service. So the person goes home and maybe does the plan and maybe doesn't. And whether that's an effective use

00226

of, whether it's really maintaining or whether it's just sort of here you go and go home.

DR. SATYA-MURTI: True. This has been very illuminating, not only to identify gaps with specific modalities and the issue at hand, but we've identified even larger gaps in terms of methodology and rigor, so some of these were brought up today. They exist with many technologies and coverage issues that come before CMS, but in this particular stroke population they just seem to be amplified even more so, is what I hear. The heterogeneity, applicability and validity of indices.

So with that, my part is over, I think.

Louis.

DR. JACQUES: Yeah. I think my role is to say thank you, you guys did a very good job with a very difficult topic, which means we put little stars next to your names so whenever we have a similarly difficult topic, we kind of know who can deal with it, so you get more of the same. Seriously, thank you very much, and thank you to the members of the public in our audience.

(The meeting adjourned at 3:19 p.m.)