AMERICAN STATISTICAL ASSOCIATION

COMMITTEE ON ENERGY STATISTICS

MEETING

October 26, 2001

The committee met in Room 6E069, 1000 Independence Avenue, S.W., Washington, D.C., at 8:30 a.m., Carol A. Gotway Crawford, Ph.D., Chair, presiding.

PRESENT:

Carol A. Gotway Crawford, Ph.D., Chair
Jay Breidt, Vice Chair
Mark Bernstein, Member
Mark Burton, Member
Thomas Cowin, Member
James Hammitt, Member
Nicholas Hengartner, Member
Randy R. Sitter, Member
William Weinig, Member
Roy Whitmore, Member

ALSO PRESENT:

Fred Mayes, Office of Coal, Nuclear, Electric and Alternative Fuels
Renee Miller, EIA/SMG
James R. Knaub, EIA/CNEAF
Jay Casselberry, EIA/SMG
Orhan Yildiz, Contractor

I-N-D-E-X

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

8:37 a.m.

CHAIRPERSON CRAWFORD: Good morning, everyone. I'd like to go ahead and open the meeting this morning. I have two announcements about procedures. First of all, any EIA staff, committee member, or member of the public who were not present yesterday should introduce themselves and please use the microphone for our transcriber.

MR. FREIHEIM: Hi. I'm Steve Freiheim. I'm with Edison Electric Institute, the Statistics Department. Linda Spencer used to come to these meetings representing EEI, but she just left us last week, so we just have two people in our department now and my co-worker was out sick, so we needed coverage yesterday. So I'm here today.

CHAIRPERSON CRAWFORD: Okay. Welcome.

MR. FREIHEIM: Thank you.

MR. KNOB: Jim Knob, EIA.

CHAIRPERSON CRAWFORD: Good morning. Anyone who did not sign in yesterday is asked to sign in this morning at the break or before you leave and the sign-up sheet is out there with our ASA representative, Linda Minor. Lunch for the committee will be held at conclusion of this meeting in Room 1E226 where it was yesterday. We'll use this time to suggest topics for the EIA spring meeting. So if you have any ideas for those topics, please let Bill know or Jay Breidt know or you can let me know. I'm probably going to leave right after the meeting. I have a flight out of National and I need to get there early, so if you have something that you want to suggest to me, maybe see me before lunch.

We might as well go ahead and get started, Fred, if you're ready. Our first presenter this morning is Fred Mayes. You should have a handout from him. Issues in Alternate Transportation Fuels.

MR. MAYES: Thank you. It really is a privilege to speak to you today. I'm with the EIA's Office of Coal, Nuclear, Electric and Alternate Fuels. When I last presented to the EIA/ASA Committee in 1984, I introduced myself as being from the Office of Coal, Nuclear, Electric and John Carlin, and that's because John Carlin was the only person that worked on renewables. I'm pleased to say now we have an entire branch that works on renewables and alternate fuels and so we are trying to give these issues serious attention.

Today I'd like to talk to you about issues that we have in trying to estimate the number of alternative fuel vehicles and the amount of alternate transportation fuels that is consumed. The first thing I'd like to go over is just to talk a little bit about what they are. Basically, they are non-gasoline, non-diesel, non-petroleum based fuels. For the purposes of this discussion and the current market, basically they are propane, either the compressed or liquified form of natural gas, alcohol fuels such as E-85, that is 85 percent alcohol, 15 percent gasoline, methanol, electricity, biodiesel, and then fuels that really haven't come into the commercial market yet. Hydrogen and so-called P-2000 fuel.

What are alternative fueled vehicles? Well, simply put, they are vehicles that work on fuels other than gasoline and diesel sort of. The Energy Policy Act I'll discuss in a minute has some fine points that I am probably not going to go into but basically if you think of those fuels that you saw before with the exception of biodiesel, it's vehicles that operate on those fuels. Any diesel engine can use biodiesel, so there really isn't any such thing as a biodiesel vehicle.

We're talking about a rare event. If you look at the number of alternative fueled vehicles that there are, we're guessing -- and no one has an exact number, we're the only ones that have a complete estimate that I know of -- but they're four and a half percent of all the vehicles in the United States, on-road vehicles. An alternative fueled vehicle can be a so-called dedicated system. That is, you've got to use the alternate fuel. Or it can be where you can use one fuel or the other but not both at the same time. Or it can be duel fuel. That is, you can throw two fuels together and the engine figures out how to deal with it. A version of that, which is the most common one on the road today is the flexi-fuel vehicle where it can take any percentage from zero to 85 of an alcohol fuel.

There's a new type of vehicle coming on, the so-called hybrid electric vehicle or HEV. you may have seen or heard about the Toyota Priez or the Honda Insight. If those names are familiar, they have an electric motor onboard in addition to the gasoline motor. Unfortunately, for some purposes, the Energy Policy Act's attitude is well, if it's gasoline that's driving the electric motor, then it's not alternative fueled vehicle. But for our purposes, there's a lot of reasons why people are looking at hybrids and we think it's important so we've been including them in our work, regardless of what the Energy Policy Act has to say.

There are lots and lots of off road alternative fueled vehicles. Lots of forklifts. Airports are trying to go electric, go propane, go anything to try to cut down on emissions. In addition to that, in the on road vehicle category, there's a never ending continuum. It isn't just cars and trucks. There are these so-called low speed vehicles, neighborhood electric vehicles, that are coming out. There are electric bikes and alternative fueled vehicles go all the way up to heavy duty trucks. So we have a little concern. We'd like to cover the waterfront but, especially when you get off road, it becomes a little bit really dicey to try to find frames for those people.

There really have been two lines of history of alternative fueled vehicles. I'm going to start with the one that came out of the Energy Policy Act of 1992. It set up a reporting requirement for us which basically said you need to estimate how many AFVs are in use by vehicle type and geographic region, estimate replacement fuel consumption. Replacement fuel is simply alternate fuel. And, in addition, that portion of regular gasoline that is so-called oxygenated gasoline, MTVE or ethanol. They told us we want you to survey people who make vehicles and we want you to survey alternate fuel suppliers for the amount of fuel that they supply. So that's what EPACT said to do.

Now,. the goal -- and this I think is important to remember. The goal of EPACT stated very clearly is to reduce petroleum consumption in the transportation sector. That is its goal and the alternate fueled vehicles were a means to that end that they set up.

The Energy Policy Act as well as Executive Orders require federal, state governments and altercate fuel providers to acquire, beginning in 1998, AFVs. However, they did not require them to use alternate fuels. The reason was politically it would have never passed because there were virtually no refueling facilities that were in existence then. So it really didn't make any sense. However, that has led to some odd quirks in the market.

What's happened over the years is a couple of things. #1, clean air has become -- until September 11 -- has become the dominant force in what's been driving the alternative fueled vehicle market. In fact, if you look at our forecast, you'll see that over the next few years there's a big wrap up and it's mainly due to vehicles that are going to be acquired because of California's zero emission vehicle mandate. But because of that, we got interested in heavy duty vehicles, even though the Energy Policy Act

EPACT history of alternate fuels was a market-based system that started 20 - 25 years ago or more, and that is people were converting vehicles to operate on propane for economic reasons. It was technologically very easy and very inexpensive to convert a truck and even a pickup truck and just run it on propane. People did this in municipal governments and in private fleets quite legally. They also did it, we believe although we don't have proof of this in the numbers, illegally in the agricultural sector because since they had access to bottled gas and that bottled gas didn't have the road tax on it, voila, they had a really cheap deal.

So propane vehicles were big before EPACT was ever thought of, but in 1997 that all came to a crashing halt because the Environmental Protection Agency issued what they called Memorandum 1A which basically said you can't certify just the conversion kit. You've got to certify the conversion kit installed on the vehicle. And that became a much more expensive proposition, and conversions, as our data show, just went in the tank. It went from 70 percent the hoses and upgrade them, spend an extra $75 and voila, we've got E-85 compatible agent. Out the door.

So what you see in this graph over here is that almost everything that's being made now, 95 percent of the vehicles made in 2000, were E-85 vehicles. Most of those go into pickup trucks, their vans, their autos, and they go people who neither know nor care that they have an alternative fueled vehicle. So we went from an environment where the AFEs were made just for people who were going to use them as AFVs -- that was a nice controlled environment for us -- to now the ones that are made available go everywhere and even though we really only care about those that might use alternate fuel, it creates an additional problem in trying to track them and figure out who's using them, who's using them in fleets where they might stand a chance of actually using alternate fuels.

I won't go through the vehicle consumption slide except to say that if you did some math and compared the vehicles and the fuels, you'd discover that LPG has a greater share of the alternate fuel than it does the alternate vehicles. That's because LPG is almost exclusively used in low miles per gallon trucks.

Then finally, the little picture of the Untied States over there shows where they're used and a fourth of them are in California and Texas. Texas, being different in every other way, it's no exception. They're different in alternate fuels. They have lots of propane vehicles and they use lots and lots and lots of fuel in those vehicles. So they are real alternate fuel consumption buffs.

There's three references if you want to get a quick overview. I don't have the ability to put those pictures up. I apologize, but I brought them in case anybody was really interested. I can put them on an overhead. But those are sort of summaries of our data.

Let me go on to what the challenges are that we face in getting good alternate fuel information. So far as the vehicles go, AFVs can be made by both OEMs and converters. There aren't many being converted now, but one can envision a future from, which is my last point up here, that it's difficult to locate those old propane vehicles because most of them were conversions and they were done in onesies and twosies and threesies sort of all over.

I mentioned what happened to the flexi-fuel vehicles beginning in 1998 and that creates a challenge because these vehicles are not just going to AFV or alternate fuel using communities. They are going to the public at large. These are stock production vehicles. If you buy a certain van, it's an '85 van and we have no option.

Dedicated AFVs. That is, those that work on one fuel only, one alternate fuel only, are a little easier because they're only in fleets but that's the good news. The bad news is there are a lot of fleets in this country and we spent -- we being EIA through our Office of Energy Market and Use -- spent a considerable effort to try to do some case studies, one of them in Atlanta, Georgia. I forget how many fleets they estimated. They started with the Yellow Pages and whoever. It was a big effort just to get the fleets in Atlanta, Georgia, that is, those with 10 or more vehicles. So just limiting it to fleets helped, but you still have some work to do after that.

In terms of the alternate fuels, it's even more of a mess because, unlike in Canada where the alternate fuels, even though they aren't used very much, are dispensed through traditional retailing facilities, refueling is often done here at private or what I call atypical facilities. That is, it may be open to the public but it's drive around the corner and at the gravel road drive in and you'll see this little man and tell him what you want and he'll point you to where the compressor is. So even that might be overcome if you said well, we'll just ask the fuel providers. They can tell you how much they've spent. Wrong, because what happens is that the fuel providers often, since they don't sell the fuel to themselves, if they're consuming it themselves, they don't sell it to themselves, it's very expensive for them to meter. The metering device for natural gas is very expensive. So if it's only one-one hundredth of the fuel coming through your facility that you put into your 900 Washington Gaslight trucks, are you really going to meter that? Well, a lot of them don't so they don't know how much they use themselves and, furthermore, they don't know how much other people use. They sell gas to Vaughn's Super Market in California or Giant Food here and all they know is they sell Giant X amount of gas and they don't know that Giant has seven CNG trucks and they have their own compression facility. They don't know how much they sell there. So we have a lot of what we call non-purchased fuel and that creates a problem for us in trying to ask either the users or the sellers of the fuel how much they have.

So that's sort of the industry, the mark, the status of the situation. What are we actually doing? Each year we have a survey of all the vehicles that suppliers make available. OEMs, conversions. We also survey certain sectors for the fleets that use AFVs. The sectors that we look are the federal/state governments, we look at them all, the alternate fuel providers which is a mixture of universal sampling, that is electric utilities, natural gas utilities and propane distributors, bus operators because A) it's a very big concern environmentally and B) because we get good data from the Federal Transit Administration and the American Public Transit Administration put together a frame and then we also look at school buses which is a little dicier.

If you believe our estimates in the entire population of in use estimates, which is an estimate and not from a survey, if you look at our survey, it looks like we're capturing about a fourth of all the fleet AFVs. We estimate the entire population of AFVs using secondary source information, and I'll say more about the problems with that later, but that's how we do that.

Replacement fuel consumption, that is the alternate fuels part of it, we use the same approach as we do to estimate the entire population of AFVs in use. I put that as an attachment in the paper, and you'll notice that there were a lot of assumptions there. A whole lot of assumptions.

So what are the problems that we have? Well, it's hard to locate all the converters. Recently there's been under four percent of all AFVs made but who knows? At one time there were 70 percent. Who knows what'll happen in the future?

The definition of an AFV is at issue. We could stick very, very closely to what the Energy Policy Act said. We've chosen to broaden it. Maybe we should call it something other than AFVs. People are interested in hybrid electric vehicles. People are interested in some of these other things, and we're responding to that information need. We just did a whole set of customer focus groups to find out what people want.

We're only looking at sectors of the end use market, and that's because we don't know where to go to find lists at any reasonable cost to use to look at the whole sector. In our estimation technique, we developed this estimate end use vehicles and ATF consumption based on -- assumptions data that is sometimes available one year, sometimes not, and some of it conflicts with our survey data because our survey data is only in certain sectors and, while it's useful to know all the federal vehicles and where they are, it doesn't help you make an estimate of all the

AFVs that are in Nevada, say, because you don't have a magical estimate that says ah, I know all the non-federal vehicles in Nevada. It doesn't work that way. You may know all the propane vehicles in Nevada or you may know all of the CNG vehicles in Texas.

So you have these parts that don't fit. One of the things that we've asked is maybe we should do away with that and bite the bullet and really just go gung-ho on a survey approach that just spends our efforts trying to look for the complete frame of AFVs.

So the questions that we have for the committee are, I've already talked about the first one really. Should we stay on this path where we're limping along trying to serve two masters or should we just go with the process of survey and, if we do, what's the frame? Where do we get the frame?

My wife put a cartoon out for me last night and she didn't think that it had anything to do with my presentation this morning, but it's really apropos to this whole problem of where you find the AFVs. It was Gilbert and it shows the guy giving a demonstration. It says, over here we have our random number generator and the random number generator is speaking and it says, 99, 99, 99. And so Dilbert says, are you sure that's random? And the answer is, that's the problem. Randomness. You can never be sure. And that's the problem we have. We have no idea. We do not know when we've got all of them or where to find them. So that's sort of the dilemma we're in.

What should we use for the target population to collect the various pieces of fuel and transportation data that are required? One option is to look at the census vehicle inventory survey. They survey. For some reason, it's only for trucks. I don't know the history. I worked there 10 years and never came in contact with the survey, so I don't know what the history was. But it only looks at trucks which fortunately start all the way down at vans and pickups but excludes cars. It also excludes buses. So we're going to be meeting with them Monday to try to see, you know, maybe they could draw us a sample of just an AFV only sample. But even if they did, we're still missing cars and so I don't know what to look for.

We've thought about R.L. Polk and looking at some of their data. If you've had good experience with them in the vehicle area, I'd like to hear about it. I looked at it mostly fuel codes and gas and diesel were blank. So that didn't help a whole lot. Or maybe we should just try to expand the '86 survey. We could probably get a good estimate of government vehicles because the census people have the survey of governments. We could probably get a frame of government people getting a good survey of municipal governments and we could at least cover the whole government front. But that leaves out the rather large private sector.

Maybe we could sweet talk suppliers into furnishing geographic vehicle type information. Haven't been able to so far but we haven't talked to them in a few years. That would at least let us know where they went. However, eventually those vehicles are going to go out of service and we will never know by doing that when they go out of service whereas if we ask the user whether he has the vehicle or not, at least you know whether or not he's using it.

Another question here and that is if we do go into the sample business, if that's your thought, are there any special methods we should use because we're going to be faced with what I call lumpy or rarely occurring events. Obviously if you're surveying from a vehicle identification number frame, it's a rare event because you're only talking about one vehicle. If you're talking about surveying fleets, it's likely to be quite lumpy because most fleets either have a bunch or they don't have any at all. So your chances of hitting a fleet that has AFVs if you're in one of those areas that we survey, is pretty small. But when we do hit them, somebody like Schwann's, the people that have the frozen food, they don't disclose the exact number but they run 3,000+ propane vehicles. So when you hit somebody, the number can be fairly large.

Should we attempt to collect alternate transportation fuel consumption information directly or maybe it would be better, instead of all these problems -- oh, I forgot to mention one problem here is that the pumps, even the traditional fueling pump, has a set of computer code that is nation-wide that's been used for years that does not allow for distinction of alternate transportation fuels. So even some of the companies themselves have clever schemes like well, if it's coded diesel but it's a light duty vehicle, we know that in our fleet that means it's ethanol. So there are some real problems in collecting the data directly.

So maybe what we should do is ask them instead things like how many vehicles do you have? What's the average milage you drive the vehicles? What percentage of time do they run on alternate fuels? A few other things. And let us estimate rather than just get the lack of information that we've gotten so far on our survey.

And finally, this is a little more sort of a policy or an EIAism, but there are some statistical issues. If you think that we ought to be collecting information on off-road vehicles and electric bikes and everything, there are real issues of where the frame should come from, so we put on as sort of the last question, what should the scope of our AFVs be? We've defined it pretty much to be heavy duty and light duty on-road vehicles for the EPACT categories plus a couple of other technological categories like hybrid electrics that EPACT doesn't recognize.

So that's it. Thank you for the opportunity and I'll be glad to entertain questions and do the best I can to answer them.

CHAIRPERSON CRAWFORD: Actually, right now I think I'd like to hear from our discussants and then we'll take the questions. Thank you. Our first discussant is Mark Bernstein.

MR. BERNSTEIN: And I will talk into the microphone today so you don't have to yell at me. Thank you very much for a really good presentation. They're not always so lively and particularly first thing in the morning, it's useful so thanks.

When Carol asked me to do this, I had read the title so I said yes, I know lots about alternative fuels and alternative fueled vehicles. I'd be happy to be a discussant. But what I don't know is anything about statistics and survey sampling and things like that. So I'm not going to be able to give you those types of answers. Hopefully Roy will be able to because he's following me. But what I can do is try to ask a few questions that sort of may lead you to some of the answers that you want to get and then maybe bring up some other suggestions.

The first question is why do we want this data and the answer is not because of EPACT. So be thinking about how should we really be using this data? In my view, the data is important to understand the impact of government policies on promoting alternative fuels as well as to get information out so people might think of doing alternative fuels, which tells me that I don't really care about converters because government policies are moving that and people are not going to be swayed by the fact that we know how many people are doing converting. So I personally don't care about how many people convert and I really don't also care about pre-EPACT vehicles.

MR. MAYES: I'm sorry?

MR. BERNSTEIN: I don't care about pre-EPACT vehicles. They just seem too old. But just going back fundamentally and thinking about what we would like to use this data for may help you figure out what to address.

I do think trying to get information from counties and cities is important. You do mention trying to get to municipalities, because I believe that that's a relatively large and growing source, particularly in California, but I also think other places. And there's certainly the Conference of Mayors and places like that that one may be able to go to. The other place is I assume you've talked to Clean Cities Program, the Office of Transportation Technology. They may have better access to helping you figure out who's at least signed up since they've been doing the program and get started there.

I guess the other place to try to think about fleets are trade associations about trying to get trade association help to work with their members to find out who has alternative vehicles in their fleets because a lot of times we find in other areas trade associations are very interested in helping find out what their members are doing to promote certain things, and right now the issue of energy security, national security, and I bet you you'd find a willing audience and a lot of trade associations to survey their members and find out who's doing what to reduce energy use.

The issue of scope. I would say don't limit from what you have now but I certainly wouldn't go to scooters. I'm not sure that's worth it. The golf carts and think vehicles and things like that. Companies that are beginning to do that, particularly in California, are going to promote how many they sold so there should be a pretty willing audience that way.

So that's about all I can really say on these things.

CHAIRPERSON CRAWFORD: Thank you. Roy.

MR. WHITMORE: I tried to prepare essentially remarks in four areas. First, estimates, numbers of alternate fuel vehicles in use, estimates of amount of fuel used, and the question of using surveys and finally the question of scope. I think you're dealing with some very complex problems here in terms of being able to survey these populations, and

I think I may have possibly missed -- I'm sure I missed some of the details. If some of my remarks are not exactly on target, just let me know.

With regard to estimating numbers of on-road alternate fuel vehicles that are in use, my understanding was that your goal is to be able to estimate that by four or five different types of fuel and by geographic region. I essentially got two types of places that you might go to for that information. You've got the manufacturers and converters and you've got the actual users of alternate fuel vehicles.

At this point, it sounds like you're not getting a lot of information from manufacturers and converters, and that's possibly a very source of information that you need to try to capitalize on and do more with, if possible.

I'm not sure I got this right, but my understanding was that there is an EIA 886 form, the made available portion of the EIA 886 that's currently going to manufacturers and converters to the extent that you can find converters, and I believe what you're getting is number of vehicles that were delivered by type of fuel that they're made for and geographic region. The paper said, and I wasn't sure I got this right because it sounds ambitious, that you thought the manufacturers and converters could also provide the numbers in use by type and region, but they're not willing or haven't been able to get that information so far. If you could get that, it would seem like that that would be very valuable to be able to get that directly from the manufacturers and converters. It doesn't seem like an overly burdensome thing to ask them for if they have that information. I encourage you to try to figure out a way to get it from them.

Ideally, getting into even more owners's use of their records, if there was some way to actually get lists of alternate fuel vehicle owners to use as sampling frames, that would be an ideal way to get at vehicle users. At a minimum, if it's not possible to get all of their information on who owns alternate fuel vehicles, if they might be able to provide information on private fleets that are purchasing alternate fuel vehicles.

All of that is probably getting into confidential business information and, therefore, would be somewhat difficult. I'll just throw out as a possibility that the federal government is really interested in getting good information on these sorts of things. Maybe there's some way of vamping some legislation that requires them to provide certain information for use as sampling frames, at the same time ensuring confidentiality and restricting the way that information could be used. So maybe in the long run you can develop ways of getting more information from the manufacturers and the converters.

You talked about the difficulty of creating frames of converters. It may be that you could periodically look at certain geographic regions for looking for better frame information on converters, especially if it kind of comes back into vogue. As you were saying, it's done much today so maybe it's not a big issue today.

In terms of getting information on vehicles in use, there's the in use portion of the EIA 886. My understanding is it currently covers about 20

- 25 percent of the population so the major problem with that is it's not covering the total population. As I understood it, it's mainly missing the local government fleets and private fleets and individual consumers. The idea of adding the local government fleets to the EIA 886 or some comparable survey I think is a very good idea. That should be pursued.

As Mark mentioned, it would be a good idea to find some way of covering the private fleets, either through getting information from the trade associations or getting information on private fleet buyers directly from the manufacturers or looking, as you were suggesting, in maybe individual geographic areas, one area one year, another area another year, to try to build up the frame of fleets from other kinds of sources that are more kind of intensive in local areas.

In terms of the Census Bureau's vehicle inventory utilization survey, as you indicated, it's not annual and you need to make annual reports. It's only doing on-road trucks, so it's missing a large part of the population and it is a survey, not of alternative fueled vehicles but all vehicles, and the alternative fueled vehicles are a very small portion of the population. So unless there's some way of making that survey much more efficient, it doesn't seem like a real good way to go but certainly it has the potential of giving complete coverage to the population. So you need to pursue if there is some way of making that sort of thing more efficient.

You had talked in the paper about developing a new survey of users of alternate fuel vehicles. At the present time, it sounds like the EIA 886 is getting a portion of the users. Adding the local governments picks up another piece of the users. It would seem to me like as long as you're dealing with fleets, something like EIA 886 that gets aggregate data for entire fleets all at once is a better way to go than the individual users. When and if you get a frame available where you go to the individual consumers, then a different type of survey would be needed for those individual consumers.

As hard a problem as the vehicles are, it sounds like the fuel use is even a bigger problem. My understanding is you're trying to estimate annual consumption of fuel for use as an alternate transportation fuel, again, by fuel types and by region, and you've got similar kind of potential sources of information, sellers and distributors of fuel, and the users themselves. It occurred to me that the best source of information might vary or might differ by type of fuel.

MR. MAYES: That's true.

MR. WHITMORE: For example, for alcohol, methanol and methane, if you're interested in just that portion, those are often mixed with gasoline. If you're just interested in the alcohol, methanol and methane volume itself, if we're just talking about an annual estimate, you can get the annual amount of those fuels sold for the purpose of being used for ATF fuel, that that might be a reasonably good estimate of the amount being used in a 12 month period.

For electricity, you almost have to get end user information, either from the fleets or whatever, getting miles driven by electric vehicles which I think you're partially covering now at the EIA

886 that's going to utilities and such.

The compressed natural gas and propane seems to be a bigger problems that's covered by the EIA 886 but, as you indicated, sometimes they've got large tanks and some of it's going into tanks and some of it's going into vehicles. That's sort of a situation where they can't tell you how much is going into those vehicles and it sounds like you're stuck with some kind of an estimation problem where you're saying that the number of vehicle miles driven with alternative fueled vehicles and percent of time that they were actually using an alternative transportation fuel. Some of these fleets have tanks that are used only for the purpose of fueling their vehicles, and you can get how much they purchased during a 12 month period for use in their vehicles. That should be a pretty good surrogate in those cases of how much has been used, especially if you also have the vehicle miles driven and that sort of thing to use along with that.

You asked about whether you should be making estimates or doing a survey. Of course, in both cases you're looking at estimates that a survey, if you've got a good frame and you've figured out some way to overcome these measurement issues, then you get a way of getting robust estimates of precision and uncertainty when you do the survey with the current kind of estimation procedures where you're having to cobble together lots of different kinds of disparate information. You don't really have any way to quantify uncertainty is the problem. Until you get good sampling frames, you don't really have any choice.

As a long term goal, you should try to develop sampling frames that would allow you to do direct survey based estimates, but it doesn't look like you're close to being able to do that right now. In terms of expanding the scope of the study to include off-road vehicles and low speed vehicles, I would think you'd want to invest your resources into being able to do a really good job of what's mandated first before tackling those things.

CHAIRPERSON CRAWFORD: Thank you, Roy.

Fred, do you have some more comments and questions now?

MR. MAYES: I do have a couple in response to your suggestions. One suggestion was that we contact trade associations. We've done that. Some of them are helpful. We contacted NAFA, National Association of Fleet Administrators, because they represent municipal and privates, and somehow or other there was a little misunderstanding. We got a list from them. Then somebody in the association discovered that we got a list from them and contacted us and said if we use it, they'd sue us. This is the same organization that does their own survey of alternate fueled vehicles. They publish it every year. The response rate is eight percent. Our response rate has varied as low as 72 and as high as 100. That is not stopping us from trying again because when we talked to them last was 1997 and now there's a different sort of a milieu in the country. So we will continue to work with trade associations.

MR. BERNSTEIN: Before you move on, rather than going to the quote/unquote "fleet operator associations," I would go to the grocery, steel manufacturer associations, not the fleet operators but the company associations themselves and let them survey their members about how they do their businesses. Retail operations and things like that. I think you'll get a different response because the fleet operator association has a particular sort of agenda. These guys have different sets of agenda and want to show that they're doing something good for the country. I would approach it a different way.

MR. MAYES: Okay. that's a good suggestion.

The other comment I was going to make was both of you have indicated a definite interest in municipal fleets and I certainly concur with that and whatever we do, we will try to include those. I just want to make one point though and that is that until the day comes when we cover everything, because of our mandate from EPACT to provide an estimate of all vehicles nation-wide, we can't drop our estimation procedure because we add up the pieces one way, state, federal, municipal governments and privates, that's one way, but that isn't the way the estimation procedure works because of what's available. So it could definitely improve our survey coverage, and we will try and do that. But it wouldn't necessarily reduce too much our burden on this estimation program which we would frankly like to rid ourselves of if we can.

CHAIRPERSON CRAWFORD: Thank you.

MR. BURTON: I understand that the DMV records are in many states confidential, but there may be another potential source of gaining information through the states. Not all but most states require annual vehicle inspections. Would it be possible to modify those inspections in a way to indicate alternative fueled vehicles?

MR. MAYES: That's a good suggestion.

MR. BURTON: And in addition to identifying the vehicles, those annual inspections almost always include an odometer reading, so you'd also get a measure of usage from year to year.

MR. MAYES: That's true.

MR. BURTON: It would be a lot of work to deal with the individual states to do that, but it doesn't sound like there's going to be any easy way to do what you need done.

MR. MAYES: Okay. And I would think that in order to inspect, you'd have to know what fuel type it is.

MR. BURTON: That's right. And even to the extent that if they're offered exemptions from the inspection process or from emissions --

MR. MAYES: You'd at least have a record of exemptions. Okay. Thank you.

CHAIRPERSON CRAWFORD: For a variety of reasons, we need to both stay on schedule and divert from schedule. Right now I'd like to take a few minutes to acknowledge Tom Cowin. This is his last meeting for the committee. So I just wanted to congratulate Tom for his service to the committee, recognize the fact that he's always been one that's quick to volunteer to be a discussant. He's always been very prepared, and I very much appreciate your work on this committee, as do all the committee members. So on behalf of ASA and this committee, we appreciate your service. Thank you very much.

And now I'd like to turn it over to Mary Hutzler and Nancy Kirkendall who will also have some words for Tom.

MS. HUTZLER: Tom, I wanted to thank you personally for being on the committee for the past three years. We've certainly enjoyed your participation, and I personally have enjoyed your astute questions and your thinking out of the box and participating in meetings like the new modeling and forecasting approaches yesterday.

Tom tells me he's going to retire in May, and I must tell you that I do envy you, Tom, because he's going to be able to do things that I would love to do. He's going to be traveling. He's doing ballroom dancing with his wife, Susan, and he's also going to try to improve his piano techniques. So Tom, thanks very much.

Thank you for your public service as a member of the committee.

(Applause)

MR. COWIN: It's really been a lot of fun and it's gone much too quickly. I'd like to, on my behalf, thank all of the members of the committee and the staff members of EIA and, in particular, I want to thank the EIA ambassador who certainly I think went out of his way to make me feel welcome here and I'm sure the rest of the committee members, and that's Bill Weinig. Bill, thanks a lot.

MS. HUTZLER: Tom told me last night that he really won't be missed because he's found a committee member to replace him that is an environmental statistician and he tells me she's going to do a wonderful job for us. Thank you for doing that as well, Tom.

I have the honor of presenting another certificate. This one is to Carol. Carol has been our committee chair for two years, and she is no longer going to be our chair effective the end of this year. So I wanted to thank her for her excellent job in that role. That tends to be a difficult job because she has the liaison work as Bill has for us, and so she gets to figure out what agenda items you folks can suggest to us. She also has to get you as consultants, she has to pay attention to the schedule, keep us on schedule during these meetings, and I think she's done a great job doing that. So I want to thank Carol for her role as chair and also, you're going to be on the committee for another year so we're going to appreciate that, too. Thank you, Carol.

CHAIRPERSON CRAWFORD: Thank you.

(Applause)

CHAIRPERSON CRAWFORD: So obviously through the chair they can pull one over in terms of the agenda. I wasn't expecting that. Thank you very much.

Now we'll get back to business. Our next speaker this morning I believe is Renee Miller. She's going to talk to us about measuring consumption by energy use sector and give us some options for obtaining future data.

MS. MILLER: Good morning. We have one more challenge for you, Tom, before you leave and for the committee as well. Tom has been particularly helpful in past discussions about energy use sectors. So we'd like to talk with you this morning about the challenge we're facing in measuring consumption by energy use sectors. This is a challenge that EIA has been facing for quite a while now. The situation isn't getting any better, as those of you who participated in the natural gas break-out session with us heard yesterday.

So we'd like to start a dialogue with you on considering options for the future. We see this as a long-range topic. It's kind of like a coming attraction. Something we'd like to have future discussions on.

To begin with, EIA provides a lot of data on consumption by energy use sector, and we show you some of the publications up here that feature these data, the monthly energy review, the annual energy outlook, the annual energy review. We also show these data in publications put out by the program offices like *The Natural Gas Monthly, The Electric Power Monthly, Petroleum Marketing Monthly.* We have annual publications, and there are probably other publications that I didn't even mention.

The point is we have a lot of data, highly visible, widely used. Go back to 1949, forward to

2020. So if we were to make any changes, we would have to be very careful in explaining the changes to the users and what the impacts would be.

We obtained the data on energy consumption from surveys of energy suppliers like petroleum marketers, natural gas and electric utilities, natural gas pipelines, and here you have an example of a survey that we send out to electric utilities and we ask them to report revenue and megawatt hours sold and delivered for the residential, commercial and industrial sectors, and we give them definitions. The electric utilities also have the option of providing data using their rate schedules. So if a customer, for instance, receives a commercial rate, that customer could end up in the commercial sector, regardless of whether it's a large apartment building, an office building, or a small industrial customer. So we don't have true economic sectors.

For some of our users, that's not a problem because they're interested in the rate schedules. For some of our users, it is a problem. And that's part of our problem, that we have lots of users with different needs. But the point is that the respondents have been able to fill in our boxes, residential, commercial and industrial.

Now we're entering a different type of problem. What if the energy suppliers can no longer report by energy use sector? Herb found for us a befuddled respondent or maybe it's a befuddled survey manager who's not sure what to do. But this isn't a hypothetical situation. We've heard that natural gas pipelines are having difficulties in reporting by energy use sector because it's no longer important to their business. They're delivering gas for the accounts of others and it's not important to them to keep data by residential, commercial, industrial.

So we'd like to talk to you about what to do. We'd like to talk a little about some previous work we've done in this area, the work that the Commentated Definitions Team did. We do have some surveys on energy use sectors that are related. I'd like to tell you about some recent developments and then ask for your help.

Some of the committee members, Carol and

Tom and Dan and Cal Kent, participated with us in discussions about definitions on energy use sectors, as you may remember. The problem we had was that EIA was using different definitions. We had different audiences, and also survey respondents had different record keeping systems. So the situation was what to do. So what we decided to do was we created a core definition. We took the part of the definition that all of them had in common.

So for instance, residential sector. All the definitions agreed that that consisted of living quarters for private households. And then we explained that there were differences by program area and we carefully catalogued all those differences and we created an energy use sector guide, and we have that online so users can click on that. I gave the committee members URL. I don't know if you've had a chance to look, but we have lots of details on all of the differences.

Before coming up with that solution though, we considered other possibilities. We considered definitions based on demand levels. So, for instance, if you were a small user, you would be residential. Medium user, commercial. Large user, industrial. Well, EIA wasn't quite ready to go that route and, putting apart the problem of how you would decide small, medium and large for all the different fuel types, there was concern that we would be missing the true economic activity sectors.

So we also considered, well, if we want true economic activity sectors, maybe we would have to go to the ultimate consumers. And we do have some experience with that. We did a feasibility study of collecting natural gas data for the industrial sector. We talked to you about this about a year or so ago. The Census Bureau helped us with this. We went to manufacturers and we asked them if they could report monthly data on natural gas consumption, and the goal was to get state level data.

While the manufacturers were willing and able to report monthly, to get state level data was very expensive. We couldn't afford it. So that's something to keep in mind. If you go to the end user and you want monthly state level data, the cost is going to be very high.

EIA does conduct surveys of energy users every four years. We have the residential energy consumption survey known as RECS, the commercial building energy consumption survey, and the manufacturing energy consumption survey which you've heard a little about in other presentations. These surveys don't completely cover these sectors though. The residential energy consumption survey, for instance, doesn't cover second homes or vacation homes. The commercial building energy consumption survey doesn't cover structures that aren't buildings. For instance, street and highway lighting is excluded. And the manufacturing energy consumption survey doesn't cover the entire industrial sector. So we wouldn't have mining and agriculture, for instance. But we do have these surveys every four years.

To summarize the data that EIA has available, I think of it as a spectrum. We have data by service class representing the rate schedules on one end and data representing economic activity on the other. On the economic activity side, we have the energy user surveys, the RECS, CBES and MECS that I just mentioned, and also our mid-term forecasts which are benchmarked to those data I would consider on the economic activity side.

On the other extreme, on the service class side, we have the sales and delivery data from the suppliers so an example would be that form I showed you earlier and our short term forecasts which use those data. Somewhere in between we have the data from what I call across system reports or integrated statistics reports like the monthly energy review and the annual energy review. They're probably closer to the service class side because they do use data based on the sales and deliveries forms. But in cases where they know about discrepancies and they can adjust, like in the electric power sector, they include street and highway lighting in the commercial sector. So they do try to make some adjustments which is why I put them somewhere in the middle.

So now you have an idea of what we have and you may be thinking well, what else do I need to know about this? So some recent developments in the electric power sector. In the past, we've been showing data for electric utilities but now we have independent power producers and co-generators as well. We've been collecting data from them for quite a while, but we haven't been consistent in how we've been reporting them. In some cases, the independent power producers and co-generators are in the electric power sector.

In other cases, they're in the industrial sector. So this is something that we're working on. I think we have consensus that the independent power producers belong in the electric power sector. The situation with co-generators isn't as straightforward. They're manufacturers. Arguments have been made they belong in the industrial sector. Another aspect of this problem is that the co-generators, in addition to producing electricity, produce useful thermal output. So the question is how does one separate the consumption out? This is something we're working on. It's part of the electric power project that Mary had mentioned yesterday in her opening remarks. So we're looking into these issues, but we're not quite there yet.

Something else to consider in the natural gas area. We've been finding that marketers don't keep their records by traditional end use sectors and now we've been finding the pipelines don't either. Something else to think about, not really a recent development. This is something we had mentioned to the committee a while back. Seymour Sudman, in thinking about this problem, pointed out that maybe what we're going to have to end up doing is some sort of a messy combination. Right now we have supplier surveys on the one hand. We have end user surveys. We have forecasts. We think of these as distinct things. But maybe the way to solve the problem is to somehow combine all these.

At the time, we liked the expression messy combinations, but we really weren't ready, I don't think, to go that route. That may be the way to go. Now may be the time to think about that. And so now it's your turn. We welcome your suggestions and ideas on next steps. Thank you.

CHAIRPERSON CRAWFORD: Thank you, Renee.

Are there any comments or questions?

MR. BERNSTEIN: I'll start and I'm sure that'll lead other people to jump in.

I have a couple of comments. I certainly understand the problem. As we talked about yesterday, since the market is changing, it's not clear where to get information. I will say that I think the natural gas suppliers you're going to are giving you a load of bull, and I'll tell you why because I work with them, too. They do keep records by customer class because the load profile is different. What they may not be doing traditionally is the people who you're talking to, they don't care about it but there are the folks who do the buying and selling and the trading end, they care about it because they need to know how much they need at certain times, and there are clearly seasonal profile differences between residential customers, commercial customers and industrial customers. An so somewhere in those organizations they do. How you get at that to convince them that they do, I'm not exactly sure but maybe we can have a dialogue about that later and try to sort of get to some of these folks and find out where in their organizations they do. But I know they think about it because at some companies I talk to are trying to talk about doing weather hedging because their residential customer is a different profile. That's one thing.

My first reaction from reading the paper was that we should get some legislation and require energy suppliers to provide data. I guess I still buy into that concept. I think that because we're changing the market place, that doesn't mean we shouldn't require these companies to provide information the way we want, and I think maybe we need some legislative action or maybe an executive order or whatever it takes to require some of these things to continue to happen because I think they're important.

I was kind of curious why you had IPPs up there for end use issues. I understand the co-generation side of it, but I'm not sure why IPPs. If we're talking about direct access, retail access, that's different but you still have the local utility who can provide the basic use data, so I'm not sure why you care about IPPs at the end use side.

MS. MILLER: Well, we were thinking of them in terms of end users. They consume energy and so what sector do we put them int? We have them, for instance, in industrial sector for natural gas in some of our publications and in some the electric power sector.

MR. BERNSTEIN: I guess maybe we got to change the definition of utilities from the conventional utilities to electricity generators and electricity deliverers or whatever.

MS. MILLER: Right.

MR. BERNSTEIN: I realize it's going to really continue to be difficult to do this, but it does continue to be important, I believe, at the sectoral level and continuing to maintain that. One of the things we're working with the department is on issues of energy intensity and how to set goals with the National Energy Plan and energy intensity. If you don't have information at the sectoral level, it doesn't make a whole lot of sense to have a national goal for total energy intensity. So if we're going to continue to track and measure as we go out into the future, we need the sectoral level information and we need it at a disaggregate level at the state level because there's a lot of variation amongst them. And it needs to be consistent with the historical record because we've got to be able to trend. I know that's going to be very difficult, but I think it's an important critical aspect for satisfying the national energy market.

MR. BURTON: I just want to echo what Mark said, and something that we said yesterday in the break-out session I think is worth repeating. That is as the nature of these industries change, it becomes even more critical from a monitoring standpoint that this information be available.

I do a great deal of work in transportation and telecommunications. Compulsory reporting is not an unusual thing, and so I think the notion that we find some of these things that we think are important to monitoring the evolution of these industries and simply require, do what is legislatively necessary to require the continued or additional reporting. Certainly at any point if it becomes useful to have an economic justification for that, I'd be more than happy to provide it.

CHAIRPERSON CRAWFORD: Any other questions or comments from the committee or Renee? Thank you very much.

Our next speaker Knaub and he'll be updating us or clarifying small area/imputation technique and presenting the results from a study of bias.

MR. KNAUB: I couldn't manage to get my CD up on the laptop, so we're going with a down grade in technology and my back-up viewgraphs. This is a follow-up to a talk I did at the fall meeting in 2000, a year ago, covering two topics. The first is to clarify the small area nature of the methodology and the second is the committee had expressed concern about bias due to model failure. Although I have studied accuracy extensively before this, I've now specifically looked at bias and will show some results.

The joint statistical meetings that we've recently had, the paper from that, acknowledges the committee for a push on studying bias. Every time I turn around, there they are. Can't play right now. Sorry.

CHAIRPERSON CRAWFORD: That's okay. Who are they?

MR. KNAUB: My dogs.

CHAIRPERSON CRAWFORD: What are their names?

MR. KNAUB: Harry and Bo. Maybe they should be Laurel and Hardy. If everybody got along as well as they did, I think we'd be in pretty good shape.

The first thing here, application across strata. This is an illustration of the small area method, the borrowing strength feature of this method. This is only an illustration, so the number of observations is generally too small and -- mission groups and publication groups are too large. Like talking about entire sets of states. We might not want to look at all northern states together, but for an illustration though, suppose that an estimation group would be the groups of data that we put together for purposes of applying models and suppose that we're looking at a situation where we're, say, collecting generation data from a certain type of plant and we look at northern states and southern states separately because of maybe the main consideration might be heating degree days versus cooling degree days.

So perhaps one model would apply across all northern states and a different model would apply across all southern states. The idea is to use the largest group of data for which a single model would apply. The more data the better, as long as the data are accurately modeled by the same model. Otherwise, you should break it into more groups. But suppose you had two estimation groups, e.g. the northern and the southern states, but suppose on the other hand you don't want to publish those groups. You want to publish a number for the eastern states and then publish a number for the western states.

If you think in terms of imputation, the letters there represent numbers and the observed numbers represented by the circled letters and the imputed numbers would be represented by the noncircled numbers. We were collecting the data. We want to use all the data we have across the northern states in order to impute for the observations we don't have and all the data across the southern states imputes observations we don't have there.

So for totals, that's easy. Just sum of the observed and imputed numbers that fall in that publication area. If you look at the eastern region, you just add up all the observed and imputed numbers there, regardless of how you arrived at those imputed numbers. Imputed numbers should be found using the most relevant data and, even though some of that data weren't in the region that's actually part of the publication area.

The reason why I use PG 1 equal one for east and PG 1 equal two for west is because there can be a PG 2 breakdown. Suppose you had northeast versus southwest or whatever. Once you've collected the data and you've gotten an imputed number for everybody, you can regroup it for totals any way you want.

Variance estimates are a little bit more involved. Next please. Still the same picture there but a couple of examples. The first example is for a total. That's easy enough to picture. Seeing which data points belong there. But for the estimation of variance, we have to look at each of the different strata that are involved so that, for example, if we wanted to publish the western region and we wanted to know what the variance was for that, we've got variance of the prediction error for use of the imputed numbers. Look, for example, at the southern strata of the western publication area. This part right here. K and S are numbers that we had to impute. Well, there's variance information that I use to carry along with each of those numbers. The variance of the prediction error is part of it. But then you also have to consider the variance due to the coefficients, variance of the coefficients in the model. So there's a little bit of extra information to carry along with those. So for each imputed number, they associated various information that has to be combined and that's -- paper has been in the past and some references show you how to get into that.

In order to be able to get a variance estimate, you have to get each of the different strata, one at a time, the southern strata for the western region and then separately the northern strata for the western region. Next one, please. Oh well, actually, stop a second. First I wanted to see if there were any questions about that before I go on to talk about bias.

Looking at bias, the models are involved here so there is a potential for model failure and something to keep an eye on. Here we have an example of where I used over 200 sets of test data. The test data were generation data by state, by fuel type and by utility or non-utility. Data were removed from these test sets to form artificial samples and using artificial cutoff and then the method was applied and then the usual Z values are calculated. The estimated total minus the observed total divided by the standard error of the estimate of total.

You can see that this graph is not symmetrical so there is bias. If you look at the median value in there, I think there were something like 224 cases. The Z value is about .3. Next slide, please.

This is the same slide again but a little bit of extra information included. About a third of the samples had absolute values of Z greater than one so, therefore, at least you're in the ball park of being able to say your accuracy is that you're close to two-thirds of a time you'd be within one standard error. The failures are too thick though and it is lop-sided there. Next one, please.

So if you subtracted three-tenths of a Z there which is equivalent to subtracting three-tenths of a standard error from the estimate of the total. That's my value over there along the Y axis. That would fix the bias in a rough sense. Next viewgraph, please. This is the same viewgraph again but with a little more information about it. The tails are still heavy of course because nothing has been done about that there. Now, the question is perhaps this should be customized by fuel type, and I have an investigation about that. I'll show a little bit of that coming up. On the other hand, in the other direction, what if it is a case of over-specifying, just like over-specifying any model.

Let's look at breaking it down a little bit though. Next viewgraph, please. I did further investigation with just utility data because that was something I had better access to at the time plus utility data appeared to generally have less non-sampling error than the often much smaller non-utilities have in their data. Still, even though these are in general larger than non-utilities and generally have less non-sampling error, I would strongly suspect that those two very small numbers, the Z values that are -6.13 and -4.38 are liable to be unresolved non-sampling error. I like to guard against that kind of thing. Next viewgraph, please.

Gamma is something I mention in my papers. It's a parameter that's long in use. It deals with heterodyscicity and it's found in regression in establishment survey data. I found that gamma = 0.5, which gives you the ratio estimate, is very robust for estimating totals. It's very robust for estimating totals. It also slightly over-estimates the variance here. Ken Brewer has an interesting discussion of gamma values in an upcoming book tentatively titled "Sampling Basu's Elephants, Combining Design Based and Model Based Inference." Publisher is Arnold Lyndon. So I've got some reference to that if you all are interested.

So I used gamma = 0.5 for everything except for natural gas in this graph here. 0.8 is a good number a lot of times with a lot of these data. It's just that 0.5 is a little more robust. But natural gas looked like it would work a little bit better with that parameter. So doing that, I'm not using much of any adjusting. It's just pretty much straightforward. But suppose you wanted to tweak some. Let's look at the next viewgraph, please.

If you wanted to tweak it a little bit, suppose you looked at hydroelectric generation and you changed gamma a little bit. Delta is just a parameter from the fact that with my formula for estimating total variance. We start with individual variances in order to help account for the variance due to the coefficients. Delta is not a particularly sensitive parameter. So 0.3 works pretty well for that. So that's sort of constant here.

But suppose we mess with gamma, like I said. Change it to 0.7 for this particular fuel type and suppose we subtract off two-tenths of Z which is equivalent to subtracting two-tenths of a standard error from the estimate of the total. We're doing a little bit of tweaking there. Didn't completely balance it out but just to see how much that would affect the results. You can see here that well, it's nearly symmetrical and it's nearly got the right fit for the variance.

As you look at the next viewgraph, if you looked at hydroelectric generation and didn't make those adjustments, only used gamma = 0.5, then you can see that we're less likely to publish misleading estimates in that the variance is over-estimated a bit there and some of the data we've got, we've got a lot of cases where if we publish data to five digits, they figure it's accurate to five digits and they hardly ever look at any information we have about accuracy. They don't want to look in the technical notes or footnotes or anything. They just see a number and that's it. So I like to be a little on the conservative side there.

Next viewgraph. Okay. This is just another way of looking at this.

MR. HENGARTNER: Could you just put that slide back up. I'm a little confused with all these plots. You're plotting the Z value on the Y axis. For example, if I look at 21, that means there were 21 which were less or equal to that exact Z value of 25.

MR. KNAUB: I meant to mention it. I forgot to. Each of these bars represents one guy. Right. What I could do is just have one data point at the end of each of these points. It's just that it was easier to see this way, I thought.

MR. HENGARTNER: It's a histogram.

MR. KNAUB: It's not really a histogram because there's only one data point. The bars came up in a -- plot that I looked at. I said oh, that makes it easier for me to see. But I should have mentioned that when I started. I've looked at them so long, I forgot. It would really make more sense to show somebody who hadn't seen it before. There's just one point here and this is a list of points. The smallest Z value I got was this guy here.

CHAIRPERSON CRAWFORD: And your model residual is basically sorted by size.

MR. KNAUB: But sorted by magnitude.

CHAIRPERSON CRAWFORD: Sorted by magnitude.

MR. KNAUB: Right. Sorry.

MR. HENGARTNER: Okay. Because I have to count here and I was just confused.

MR. KNAUB: There are 41 guys that were in this group.

MR. HENGARTNER: I get it now.

MR. KNAUB: Thank you for helping me clarify that.

MR. HENGARTNER: Forty one hydroelectric plants.

MR. KNAUB: Right. Data from 41 different hydroelectric plants. Right.

MR. HENGARTNER: Thank you.

MR. KNAUB: Thank you. I had that in my

notes somewhere but I missed mentioning that.

So this one is the case where I did some tweaking. You can see it looks more normal except for the kind of fat tails there. This is the adjusted one. You can see that I made it look better. This is the adjusted one where I messed with the gamma value and I subtracted out some specifically for the bias.

Next viewgraph is the one where I did not make any adjustments. Didn't subtract anything from Z for the bias, didn't use anything particular for the gamma value. This is for all the data points. I'm going to ask you to turn back one more time. I'll try one more time to get myself straight on this. This is a histogram where I have combined. This is not like the previous graphs we discussed. This is an actual histogram. This is an actual histogram that I wanted to show you what happens when I look at all the data for the utilities for the different fuel types and each of the different fuel types I did a special adjustment for it using a different gamma value and in a couple of cases subtracting out something from the bias to make it look good.

Now, the next viewgraph --

CHAIRPERSON CRAWFORD: Just a point of clarification. To make it look good means you want it to be normally distributed.

MR. KNAUB: Right. To take out the bias.

CHAIRPERSON CRAWFORD: The bias and also do something with the tails or just the bias?

MR. KNAUB: Correct.

CHAIRPERSON CRAWFORD: Both. Right?

MR. KNAUB: I would hope that changing the gamma would help with the tails a little bit, but it didn't help that much. The tails are still a little fat.

The next one is a case where I just used the plain vanilla gamma values, .8 for natural gas because it seems justified that the idea here is not to make so many adjustments because for one thing, gamma values, portability is a term that's in one of my papers there and it's something I got from Ken Brewer about the gamma values. If you go from one year to another with the same data sets, the best looking gamma value may change on you more than you'd like so just like any type model where you don't want to over-specify. I wanted to go to a case where I did as little tweaking as possible. So here I just said gamma equals .5 for everything except for natural gas where it looked like 0.8 was better. Later -- and I were working with the current data and it looked like we changed gamma to .5 for natural gas, too. Right? But anyway, in this case almost everything is gamma equal .5. No other adjustments made. And you can see that I've made the scales the same for this graph and the previous one. You can see the peak is a little higher in the center so, therefore, there's less chance of publishing a really horrendous number.

Next viewgraph please.

MR. COWIN: The tails are also thinner.

CHAIRPERSON CRAWFORD: Yes.

MR. KNAUB: Yes. The scale I think whose--

MR. COWIN: The tails the same.

MR. KNAUB: Because it's so much higher in the middle, then it's going to look thinner in the ends anyway.

This one, just a quick look at that. That's just for the adjusted T and the next one is for the unadjusted. Next one, please. This is comparing the observed errors to the standard errors.

I'd like to go to the conclusions here. Viewgraph 18, please. So we could adjust for each data element based on the historical data. Customized gamma but that can be not very stable. And adjust the estimate for T based on the Z-graphs but subtracting out a little bit. But that would have to be based on the historical data. I've seen examples of that in one of our forecasting models. I've seen that it looked like it was over fit for the original data and then the accuracy got stranger looking later. So I think I'd like to avoid that. Next viewgraph, please.

I found in the last 10 or 12 years fooling with this that gamma values of 0.5. As I said, I didn't go into what that really means other than it's the ratio estimate. But it's fairly robust. Performs well for estimating for T. Here you can see that it tends to over-estimate for variance. So it's not as likely to publish really bad number although the tails can be somewhat thick, depending on how you look at it.

But as I said, gamma equals .5 is a robust estimate for T and it's even good for fairly moderate to somewhat severe data quality problems. One of the things we have to be very careful about is non-sampling error. Of course, that's another reason not to want to under-estimate your relative standard error because we don't have a whole lot of information for people about non-sampling error either.

For publication purposes, this seems to work best, I think. Now we're back to the non-histogram type things. Each one of those bars is just one point is at the end of it. This is just a quick comparison of coal, natural gas, hydroelectric and oil. We can see each one of them, the weird thing is that they all seem to be biased somewhat, T is biased somewhat large. The actual bias is very small if you look at the total, you add it all up. The total isn't very biased. But they're all biased high which is interesting because if you look at the theory, it seems like, if anything, with using this kind of

The other thing is about the bias. It seems to me I very vaguely recall this because this was a year ago, but I was in that break-out session and I think we were concerned about the cut-off sampling bias, the fact that by design you are cutting off part of the population, never looking at it.

MR. KNAUB: That's what I did here.

MR. BREIDT: So you did that in this example by taking real data sets, sampling using this cut-off sampling method and then fitting your model, predicting --

MR. KNAUB: Correct.

MR. BREIDT: Okay. And I just wonder about that because these are data sets which you've now looked at. You've looked at the entire data set. You pretty much know that the model works, so it's kind of cheating in a way as opposed to the real situation in which --

MR. KNAUB: I didn't pick the data.

MS. KIRKENDALL: He uses the new data every time it comes out. I mean this isn't an old data set he massages repeatedly. He has the new data every year that he has new data. We have a census survey every year. He has the data that he can play with every year.

MR. BREIDT: Okay. I guess I don't follow that because in this situation you've got a data set. Is it a census that you're --

MS. KIRKENDALL: Every year we have a census.

MR. BREIDT: And that's what you're using. You're using a census.

MR. KNAUB: Test data. Correct.

MR. BREIDT: Okay. I didn't realize that. I thought that was a sample which wasn't self or cut-off sample but you were substituting that with a cut-off sample.

MR. KNAUB: I chose the artificial sample to indicate that. I guess I wasn't clear on that.

MR. BREIDT: Okay. But it's an artificial sample from a census.

MR. KNAUB: Correct.

MR. BREIDT: Okay. I thought it was an artificial sample from a sample. That would be a different story. So I understand that now. Thanks.

The second question regarding bias is that you've got some indications of heavy tails in these Z values and one possible source of that heavy tailedness is under-estimation of variance or using a variance estimator that doesn't really give you a pivotal T statistic. So if you under-estimate that variance, you get something too small in the denominator, you get something too large overall. Right? So you could potentially get heavy tails.

The variance estimation procedure, again my memory on this is really vague, but I think we talked about this in the break-out session. It seemed to me that rather than computing the variance directly, the variance of that sum of prediction errors which is the difference between your estimated total and the true total, the difference of that is the sum of prediction errors over the non-sampled part of the population. If you look at the variance of that and compute it directly, there's an expression right now. If I remember correctly, you add an approximation to that which you used for computational convenience. Is that correct?

MR. KNAUB: Maybe I'm not following you either on that, but the variance for each of the individual imputed numbers is not based on just what's in the publication group. It's based on the estimation group which goes outside of that.

MR. BREIDT: Right. I understand that. But you can just write down what it is. It's a variance of, you know, it's going to be the sum of the Xs over the non-sampled population times the X transpose, X inverse weighted appropriately over the sample part because that's giving you the estimation, variability times that sum again and then plus the part that is the sum of the variances of the prediction errors which are independent of everything you've seen because you have the sample and the non-sampled population. So you can write it down but you have a variance expression which depends on this parameter delta and it seems that you just make a choice for that delta for computational convenience. Is that correct?

MR. KNAUB: Well, I'm not sure how -- if

you're going to have groups where you're picking up data that are different from the groups that we actually group in for publication, I'm not sure how else you would be able to carry the information that those imputed values are dependent upon data from outside of that group. Maybe you'd have to show me somewhere exactly what you mean by that because I don't really follow that. It seems to me that this was the only way I could be able to regroup any way we wanted to in the future. I mean if we have the data stored and we have an imputed or an observed value for each element in the frame and then we want to go back any time -- like, for example, the NERC region changed on us once not too long ago. The boundary changed. If you want to go back and re-calculate the totals for that, then you can just move over those individual observations whether they're observed or imputed or not from one group to another and then add up the total. Then you could quickly also come up with a variance estimate putting together all those pieces. But I'm not sure what you're referring to.

MR. BREIDT: Well, it just seems that the variance expression you have in what we were handed out has something in it, a delta, and I'm not sure where the delta comes from because it's not part of the model. It seems to be something that you've chosen to allow you to do this variance estimation that you're describing.

MR. KNAUB: You can't just add the variance of the prediction errors.

MR. BREIDT: I know that, but you can write down what the variance is and it's just a function of the design variables X in your model, the gamma that you've estimated from the data and then the signa squared that you've estimated from the data.

MR. KNAUB: Then if you went back and wanted to re- -- you know, like I mentioned, PG1 equaled whatever. If you had another publication group breakdown and wanted to go back in there, you'd have to go back and recalculate based on those things. Kind of keep things simple is one of the main things we try to do.

One of the biggest problems. You mentioned about the tails being thick. One of the reasons for that might be because of non-sampling error. Trying to keep things as simple as possible is one of the things I think that would help with error accuracy.

CHAIRPERSON CRAWFORD: So that's a question. How involved is recomputing that mean squared error formula?

MR. BREIDT: Is just seems -- well, I guess what I'm really wondering is is the variance expression you have in here an approximation which you've chosen for computational convenience? It seems to be.

MR. WEINIG: Would you stand at the microphone, identify yourself by name and then repeat what you just said.

MR. YILDIZ: I am Orhan Yildiz. I am a consultant who works with Jim on this to actually implement it. Delta doesn't seem to play into this as much as the weighting factor gamma.

MR. BREIDT: I guess my question isn't how does delta play into it but why is delta there at all? I mean it's not part of the model, so it came from somewhere and I think it came from an approximation used for computational simplicity.

MR. KNAUB: It's explained specifically in that paper that I referenced a year ago, 1999. Would you show me the next to the last viewgraph please. There's a specific, very simple explanation of it in a 1999 paper. InterStat. That was a reference last time. I had some slides I used for that I didn't bring with me though.

CHAIRPERSON CRAWFORD: That's okay. Isn't what Jay is saying correct? I mean regardless of how you actually determine or derive delta, that is what it's doing. Right? You provide for the purpose of ease of computations an approximation to the variance.

MR. KNAUB: Right, because of the variance in the coefficients. Right.

CHAIRPERSON CRAWFORD: I think that one of Jay's points at least is that we all notice that not only was the bias sort of being adjusted but what really struck me more than that were the tails. He was suggesting that maybe the tails were too fat to begin with because delta is not quite right. Is that correct?

MR. KNAUB: Actually, this is really pretty much a non-point in this in that gamma was the thing that really was making changes there. I looked at comparing in the cases where you had estimation groups and publication groups exactly the same. Then you could very quickly and easily compare the ordinary estimate and variance to the estimate using delta there, and it was no problem then.

MR. YILDIZ: I think you have a very valid point that the model is actually failing in certain cases. In certain types of fuels, the model is failing and we are not getting good data to get a good coefficient variance. That is creating problems. That's my understanding, that that's creating large errors which is causing the tails.

We have plans to correct that by specifying the model differently. Because we have a cut off point for utilities, the cut-off point might be large, we want to compare non-utilities whose cut-off points are really small. When you put the two together, what is missing in the utilities is actually in non-utilities lower end of the strata. And if you put those together, the model may predict better. That's our guess. And maybe attack the bias and see if we can with some fuel groups by getting a better model. And I'm wondering if that might help the bias issue.

MR. KNAUB: I don't think that has much use. That was about the non-utilities. We have a lot of problems with the non-utility part. Right? But delta is really a part of this. Yes.

CHAIRPERSON CRAWFORD: May, were you finished with your comments or did you have some more?

MR. BREIDT: Yes. Thanks.

MR. HENGARTNER: I just wanted to mention. Now I understand what these parts are. We can call them the inverse of the cumulative distribution function. That's what it is.

MR. KNAUB: Thank you.

MR. HENGARTNER: The other thing is in terms of the tails, I mean they're not that heavy. There are a couple of big values but overall, I mean now that you think of them as inverse of the normal, it doesn't seem to be that crazy.

MR. KNAUB: Especially if you use a more conservative gamma of 0.5. I think that you'll be all right on that. Makes much more difference than much of anything else. Thank you.

CHAIRPERSON CRAWFORD: We just have one more presentation to go through, so I'm thinking rather than break we'll just go ahead and finish up and then be done for the day. So I think Jay is here. Our next presentation will be by Nancy Kirkendall and Jay Casselberry, and there is a handout for that this morning that I think you have.

MS. KIRKENDALL: This is a new effort and it's not only EIA, it has to do with every federal agency. The Office of Management and Budget on September 30 published a *Federal Register* notice requiring all federal agencies to come up with the information quality guidelines that I'm going to be talking about. Evidently it's a requirement that came I think because of an EPA ruling that was based on data that were not publicly available and there's some question as to the quality of the data.

So Congress has been making several requirements of OMB. I think one of them resulted in earlier legislation on monitoring quality of information. This one is the final one. Because of that one issue, now all federal agencies have to take some steps to assure quality of their information. So we're going to talk about this particular effort.

We did have a chance to participate in the development of *The Federal Register* notice and to comment on it as it was being developed. It was issued on September 30, 2000. Within one year, every agency is supposed to have information quality standards and guidelines that have been publicly announced, revised by comment and are in final form. And we have to put our draft guidelines out in six months.

The first requirement is that we have to adopt a basic standard of quality. They defined quality to include objectivity, utility and integrity. It's supposed to be a performance goal and we have to take steps to incorporate information quality criteria into our agency dissemination practices. This is a very broad definition of its information quality, so this applies to everything that EIA does. It applies to our data, it applies to our models, our analysis, our forecasts and, of course, every one of those is going to end up with different ways that you try to assure quality. It's certainly not one size fits all. It's supposed to be established at all levels. Can very depending on timeliness and so on.

Next slide. I think they recognized that one way of assuring quality is through a review process, especially for analysis and forecasting. And so they specifically say we need to have processes for reviewing the information. Otherwise, they're just saying how important they think this is to everybody. Next slide.

And the last one is that they want every agency to develop an administrative process that allows anybody to seek and obtain friction of information. It's interesting though. This thing says, "Information disseminated by the agency that does not comply with the guidelines." So it sounds like you can only complain if you didn't follow the guidelines or you can complain, but then you're safe, even if the number was wrong, as long as you follow the guidelines.

But anyhow, we certainly need to develop an administrative procedure. Statistical agencies are pretty good about trying to assure the quality of their products. At least we try hard. Probably a lot better than some of the regulatory agencies. But this is an additional burden and our goal is to try to make it as -- you know, you want to follow the requirements but you really don't want to spend a lot of time and energy doing it. Next slide.

I bring up the Inner Agency Council on Statistical Policy is a committee that consists of the heads of the 14 largest statistical agencies. Mary Hutzler is a member of the committee. I've been attending for her because she's kind of busy these days, as the only one in the front office. It's chaired by Kathy Wahlman who's the chief statistician at the Office of Management and Budget, and I used to work for her before I came back to EIA.

This group decided that since statistical agencies are very similar in many ways, we all collect and produce information, many of us do surveys, others do forecasting, many do analysis, and we all have common approaches for assuring quality. So we decided that it would be a good idea to work together to try to come up with a common approach to satisfying this requirement. Next slide.

So I'm chairing the group and I've got members from each of about 10 agencies. The numbers increase as I harass people and say, wouldn't you really like to have somebody come and play on this team. We've had one meeting so far. We have another one in a week and a half. Next slide.

So we're going to have six meetings, I guess, before we have to have a draft because we have to have a draft of our *Federal Register* notice to be published by the end of March or first of April. So our goal is that we can come up with one *Federal Register* notice for all of us that will include a list of standards or at least topics for standards that we all will have. So that allows each agency to have a different standard because there might be a different standard needed, depending on the timeliness and detail in our products. But at least we'll agree on the topics that we'd have a standard on. WE think that we should adopt a common approach to this administrative procedure so at least that if somebody wants to make a complaint about the quality, you'd do the same thing, no matter which agency you wanted to get to.

The thing I didn't mention about this administrative procedure. There's an annual reporting requirement. We have to report annually to OMB the number of complaints we've received and what we did about it.

The other thought is that we might like to have a common format for an information quality webpage that would deal with the standard. We could put our standards on there. We could put a button on that would allow people to send us a comment according to this administrative process. So what we're doing now in our group is we're looking at standards. Go on to the next slide. Each agency, in addition then, we've got this overall process but each agency, in addition, has to do its own work on this problem. And so the National Center for Education Statistics has had standards for some time and they started a process a couple of years ago to revise their standards. So they're the ones that are the furthest along in this.

The Census Bureau and the National Science Foundation, which is where Linda CArlson went, already have draft standards. They're working towards this, too, but they have more recently started it. We have had our statistical standards in place for some time. I think they were last updated in 1989. And so we've just started an internal process to review and update our standards, and Jay will talk about that effort. He's leading it within EIA.

What the inter-agency group has decided to do is to share our standards. We're starting from the NCES standards because we think those are probably more extensive than any of the others. We're going through them. Try to decide which ones are ones that we should keep as standards that we would say all statistical agencies need standards in these areas. That would be the topic of our next meeting. And then each agency will develop standards around those in this particular process.

EIA has a proposal. Also at our next meeting, we're going to talk about the administrative procedure for these complaints. What we've come up with in EIA is that it should be on this webpage. If every agency had a webpage that had kind of a common look and feel to it with the standards that were specific to that agency and, say, a button that would allow a user to say I'd like to report a problem with this data element or to register a formal complaint, something that would make it easy for us to count them and figure out whether they were pointing out a real error, make sure that we had followed the process as we said we would.

The other thought was if we wanted a webpage like that, we could spin that effort off to another inter-agency group that's called the Fed Stats Committee. They coordinate work on websites. So if we could do something like that, we might get more milage out of it.

I'll give this to Jay to talk about what's going on in EIA.

MR. CASSELBERRY: Thank you. To address this for EIA, the senior staff of EIA created a committee and they gave us a charter and we have members from all the different offices in EIA to make sure that we get valuable input from everyone and to make sure there's sort of a common agreement on what the standards should be for EIA. We're meeting weekly and we just had our first meeting last week on Thursday and our second one was yesterday. So we're just getting our feet wet and trying to get an idea of where we're going with this.

Nancy talked about the timing. The timing is pretty tight. I mean you have to have everything, the guidelines, up on your website by the end of next fiscal year, September 30. So we're working on those things. Next.

But what the senior staff did when they asked us to take this on is they said to look at the guidelines and determine what EIA needs. So we're looking at our existing standards which pretty much have been in place since 1989. The first ones were in the early 1980s. We're looking at the standards that the other statistical agencies use to see if maybe they have some good ideas that we haven't thought of.

So we're trying to identify for the future what areas EIA should have standards in and maybe with standards we usually look at those as requirements, things that the agency must do. We may also come up with some supplementary guidelines which are sort of the preferred ways of doing things that aren't sort of hard and fast. You must do it a certain way.

We're going to form working groups to develop these standards to try and bring in some of the people within EIA who work in different areas so if we're working in systems design, we'll try and bring in some of the technology people. When we get into the data processing, we'll try and bring in some of the people who know more about how to do that to add quality to that and with dissemination also.

We also need to look at our review process because the OMB guidelines do say that you have to have a process in place to review everything, all the information you disseminate to make sure it meets your quality guidelines and we're also going to be just for our agency looking at the process we'll have for the complaints and for then reporting to OMB. So hopefully it'll be a common way across the statistical agencies and we can just piggy-back on that. Next.

When we come up with these standards, we need to get endorsement from our senior staff and from the working staff because people have to be able to understand them. Any guidelines or standards we come up with have to be doable with the resources that we have. We don't want to make them so stringent or so hard that we sort of doom ourselves to failure. And we want to just sort of be able to disseminate it and go out and work with the people.

We also need to come up with a process for these standards. We're coming up with quality standards within EIA. We also have other groups that come up with standards on the look and feel of our web. We have people that are coming up with sort of the systems, the technology, the IT standards. We need to sort of bring those things together and try and make sure that when EIA staff want to know what standards apply to me, what do I have to do in these different areas, that they're not having to look in 10 different places or they aren't missing some things that are important.

As we move along, we have to report back to senior staff, the charter tells us, and keep them up to date, keep EIA staff in the loop so they know what's going on, and meet the goals and deadlines in the OMB guidelines. And then by next September 30, we have to have our guidelines up on the website and the process in place for people to comment that the information we put out, if they have a complaint that we didn't adhere to our guidelines. Next.

Nancy mentioned the timing. Hopefully the inter-agency group that Nancy is a part of will do this for all statistical agencies, but if they can't for EIA, we're going to have to have something out for public comments by next April 1 that will tell the public what the statistical agency or EIA's guidelines for quality are and ask if they have comments. Then by next July we have to get that into OMB so they can look at it and draft and see if we're on the right track and then they'll come back to us with their comments or their suggestions and then we have to have our final guidelines up on the website by the end of the next fiscal year, really by October 1.

Then each year after that, we have to do reports to OMB on how many people complained that we didn't follow our guidelines and how we resolve those complaints. As Nancy mentioned, there are differences between the quality guidelines where we say certain things and if people want to complain about the data. There may be differences there. Next.

So we're coordinating with Nancy's team and the federal statistical agencies team and, as I mentioned, we're also coordinating in-house with some of the teams we have that come up with standards that aren't purely quality but that do impact it like how things look and feel on the web and if you're putting things up on the web, how you can easily get the things like the data accuracy discussion on sampling and non-sampling errors, things like that. The IT standards and performance measurements team that we have.

I'm going to go through this pretty quickly, btu I wanted you to have a chance to know what we've come up with and we're going to give to Nancy to take to her inter-agency group to talk about what standards in general statistical agencies should have. These are pretty much based on the EIA standards. The handout I gave you, the last three pages have this exact same list of standards but it's got down what we consider to be the purpose of each standard. So it gives you a little bit more detail other than just a title. But in the model, the forecasting and analytical model area, we need something for the acceptance of models to make sure that their design and their validation are done so the final results have quality in them. Something to bring the models into EIA to make sure they're accepted for use by EIA before we start putting out products with the model results.

Make sure the models are documented so that both internal and people on the outside who are interested in the model results can see the inner workings of the model, the archive of the model results so that after they're published, if questions come up, we can go back and replicate how we came up with results and if we use proprietary models from outside sources, we need a process in place to make sure that somehow we do some internal validation to make sure about the quality of the results of the model before we start using it in our own information products.

In the inter-agency standards, we'll just be looking at things like the federal information processing standards which set general guidelines for all the federal agencies. Codes, abbreviations, acronyms, and definitions, at least hopefully within EIA we'll be consistent with how we use terms and definitions, the same coding, scheme, and abbreviations, things like that within our products so we have consistency within there. I don't think we'll ever achieve consistency across all the federal agencies but hopefully within EIA we'll be able to do that and within an agency you can do that.

And then the Office of Management and

Budget has standards, too. They set certain categories for race and ethnicity. They define what the standard metropolitan statistical areas are and things like that. And so if we're going to have data that's released in those formats, we have to make sure we adhere to the OMB guidelines and standards.

The next set of standards are more technology. They're more systems standards, at least the first two, how we design an information collection and processing system and then how we document it for our own use to make sure that the programmers know what they're knowing and the people who are carrying on the operations, both the electronic operations and also just the every day operations of fielding a survey. And the frames maintenance, just to lay out some specifics on how the frames are maintained to make sure the data remains reliable.

And then you get into the actual collection and processing. We'll hopefully have a standard on the planning, design and testing of our information collection and processing systems. How we handle our respondent contacts, how we handle non-response, how we handle our edits, what we do to validate the data and then if we have any quality indicators that we use like response rates and things like that, we try to inform the public but we also use them to improve our survey processes over time.

And in the dissemination area which is what the user sees, we were talking about dissemination practices. We have a couple of products in EIA that have set time release and they're released to the whole public at once and you tend to think of that more like with USDA with their crop reports and some of the BLS things. But if there's something that may affect the market, you want to make sure that you have certain standards on when it's release and that everyone has access to it at the same time so there's no question of favoritism.

Revision processes and procedures basically to make sure that we aren't continually updating and revising numbers for small changes for productivity and when we do revise products they're on the web site as a revision to what was there previously. Estimation, how you do estimation and just make sure it's documented or some information is given so people can assess the quality of it and what the agency did to come up with its estimates.

The presentation. We used to have a standard on graphics. Maybe this might be the same. We might have something that's a little bit broader that talks about some standards for EIA on the way tables look, on the way graph look, and I don't know whether we would ever get into it but on text. If you're going to talk about variables increasing or decreasing, is there a certain level when you say that? I think sometimes we tend to, if it went up by .01, we'll say it increased. And sometimes that's not much of an increase.

We're also going to talk about the data accuracy presentation, which is what most of you are used to seeing in the back of our publications now which talk about the non-sampling error and sampling error and things like that and how the survey was done, the general description of the survey. We'll also come up with proposing a standard on confidentiality and data access, just so the staff at

EIA especially know what's expected with regard to handling sensitive, confidential information that we get in our surveys and if we're ever going to have any access for anyone outside of EIA, how that's established. Like when you think of the Census Bureau, they have data access right now where you can get access to microdata through their research centers and things like that for statistical agencies.

The next one follows onto that and it's when we put out data in tabular form, what we're going to do to make sure that there's nondisclosure of confidential information. And the rounding is self-explanatory and then freezing information files to make sure that if there's a question later on, we can go back to the source data and then any documentation that we put out if we do have other releases besides the releases on the web and such.

That was going through it pretty quick. I know this was brought up yesterday, but right now we're focusing on the topics that a statistical agency should have standard on and maybe the topics that EIA should have standards on if it's something specific to us, so we wanted the ASA to give us some suggestions if they had any on topics or areas that will be excellent for quality information for a statistical agency and any for EIA. Thank you.

CHAIRPERSON CRAWFORD: Thank you. Are there any questions or comments from the committee?

I have two. The first one is kind of a funny thing. The acronym for Federal Information Processing Standards, is that a standard acronym or is that something that was newly created to describe these standards?

MR. CASSELBERRY: We've been using the term FIPS since the '80s.

CHAIRPERSON CRAWFORD: Okay.

MR. CASSELBERRY: That's just what I'm used to calling them, the FIPS standards.

CHAIRPERSON CRAWFORD: Okay.

MR. HENGARTNER: You're confused with the FIPS code.

CHAIRPERSON CRAWFORD: Exactly. That was what went up in my mind.

MR. CASSELBERRY: It comes out of MIPS.

CHAIRPERSON CRAWFORD: The other question I have is I know that the regulatory agencies themselves don't have necessarily quality standards. But some of their programs have been grappling with this issue for decades. I know when I used to work for EPA Super Fund we had this whole thing about data quality objectives and contractors had to draft data quality objectives to ensure that the data and the statistical methods and most everything that you've listed here with the exception of information dissemination standards because they really didn't do any information dissemination. But everything else, they called them data quality objectives. I wonder if you've somehow tried to look to see what they've had or if you haven't, maybe you haven't found it useful.

I would just suggest that I know, for instance, when I worked on -- this was part of the Department of Energy even under the WIPPP, the Waste Isolation Pilot Plant Project for Nuclear Waste Disposal in New Mexico. This would have been '89. There was this big initiative for development of data quality objectives to scientifically evaluate that site and similarly they've had one for Yucca Mountain and they've had another initiative with cleaning up of Savannah River. So every time they try to move this nuclear waste stuff around or bury it somewhere, there's the whole initiative to develop data quality standards in the scientific evaluation of those sites.

So maybe there is something. I wouldn't suggest spending a lot of time trying to find it because it might be buried, but it might be worthwhile to see just what other people have come up with when faced with a similar situation.

MR. WHITMORE: I also have some familiarity with EPA data quality objectives process, the DQO process. It was primarily kind of a standardized process for implementing the scientific method for developing a hypothesis test, determining what kinds of sample sizes that you need for any kind of a decision making process. It certainly would have some ramifications for this. I'm not sure it's real directly applicable. But there is a textbook as well on the DQO process and using S+. I don't remember right off the top of my head who the authors are, but the DQO process and S+ or something like that that's quite good. It might be worth taking a look at.

I'm also familiar with NCES standards because I do a lot of work for them and I know one of the things that they're looking at is for surveys, they're setting essentially threshold response rate levels. If your survey achieves above a certain threshold response rate, then they're essentially not concerned with non-response bias. If the survey response rate falls below the threshold, then they're going to require lots of additional analysis to try to get a handle on non-response. They're doing the same thing in terms of item non-response, looking at individual survey items. If item non-response rate is below some threshold, then they're required to do additional analysis to try to analyze the potential for item non-response.

I was just curious as to whether or not you were thinking about something as prescribed as that? Response rates above a threshold, we're not too concerned with non-response issues below that threshold and basically what they're saying is you have to do a lot of additional analysis before you can publish the results if you're going to publish things that didn't get good response rates.

MS. KIRKENDALL: I don't think we've gotten that far into it yet. I think we're still on major topics. I guess our response rate standard does say we have to a response rate greater than something like 75 percent. I don't remember exactly, but it does prescribe a number. Actually, we complain about response rates falling but they're still pretty good. Most of our surveys are mandatory which I think helps.

MR. WHITMORE: You're in much better shape than NCES.

MS. KIRKENDALL: Yes. NCES also has a lot of these -- first they go to schools and then they go to teachers and then they go to students, and so you can have non-response at each phase so that the overall non-response can be pretty high pretty easily in a survey like that.

MR. WHITMORE: They have a lot of longitudinal follow-ups as well, so you've got the first year and then multiplied by the second year response rate. It's the same kind of scenario. You can get down to some really low response rates.

MS. KIRKENDALL: Most of ours are easy compared to theirs fortunately, so we may not need a -- standard. But we'll see.

MR. WHITMORE: I think in general the agencies may decide the response rate for statistical agency is a good thing but then the agencies will define it. They'll come up with topics for the agencies and then the agencies might have different standards amongst themselves.

MR. CASSELBERRY: You're saying procedures for maximizing response rates. It occurred to me that's a little different flavor than what they're saying NCES where you may do everything possible to maximize the response rate, but you still have a 40 response rate.

MS. KIRKENDALL: Right. We do occasionally have a survey that comes in with a low response rate. They know they have to do something more. We will be looking at that and NCES has shared its standards with us. We just haven't had a chance to really go into a lot of detail on it.

MR. WHITMORE: Is Marilyn Millan participating?

MS. KIRKENDALL: Yes. She's on my committee. The other thing I noticed in going through there is we don't have anything on reviews yet. There is a requirement to do something specifically about reviews. We thought it might be easy to handle with standards. If we don't do it with the standard, we'll have to do something else. But it has to be addressed because it is a requirement.

MR. WHITMORE: What NCES does is anything that goes out at publication has to go through the standard review process and they've always done that.

MS. KIRKENDALL: Yes. They have a horrible review process.

MR. WHITMORE: It takes months.

MS. KIRKENDALL: It's both internal and external. Everybody in the agency.

MR. WHITMORE: They call it the adjudication process. I'm not real familiar. It takes a long time. It's what they refer to as the adjudication process. Once you've got a document that you think is ready for publication, then I think it goes through a review within the particular group that it belongs to. Once that group administrator officer says yes, this is ready for publication, then it goes on to this adjudication group to review for NCES. There's usually several rounds of revision. They've gotten really, really bad. They've got publication standards in terms of what to capitalize and not to capitalize and punctuation and abbreviations and all kinds of things. Really getting to the point where there are people complaining that it seems like they're more interested in what it looks like than the content.

MS. KIRKENDALL: But they have people that look at content, too. I actually reviewed one of their documents. I was na outside reviewer. It was on customer surveys. I was kind of interested to do that. They had a number of outside reviewers and inside reviewers. They call a meeting with all reviewers and you go through your comments and discuss them and there's an agreement made whether it'll be taken, what will be done about it. They write a report that says these were the comments and this is what happened with each comment.

MR. BERNSTEIN: Do you currently have a review process?

MS. KIRKENDALL: It depends on the product. For analysis products and the forecasting products, we do a fair amount of reviewing. Nothing like MCES does. For things like your weekly petroleum data, you can't do much of a review on a regular basis. That's a review internally. They do a very careful review. It's actually the division director who signs off on that. So there's only so much you can do if you have a fast turnaround thing.

MR. BERNSTEIN: What Rand does is nothing leaves Rand without having had an internal peer process that includes people who are not either involved in the project or in that unit. Everything has to be reviewed. There are different categories of how detailed the review has to be for quality of the analysis. And so everything goes through a process that doesn't have to take a long time. Clearly, stuff that has to go out on a weekly basis, we're not talking about that. But other things, particularly analysis stuff, having a review process is very important.

MS. KIRKENDALL: We do have a Category 1 review process. I think the analysis products, at least most of them, get reviewed that way. When an office thinks a product is ready, it will be sent to other offices to review. Some offices are very good about documenting the comments that they got and what they did with them. Others are not so good. I think it was Mary's office that was particularly good at documenting those things. I suspect that other offices will take up that practice, too, now that she's got more responsibility. We also do an independent expert review and get outside review.

MR. SITTER: I just feel that a lot of the stuff that's here is there are a lot of places to go except the one place that isn't really talked about here is this requirement three except to say well, we're going to put a button on our webpage. I mean a button on our webpage is going to be a small part of requirement three, I imagine. I suspect that this is a place where you're probably going to have less places to go in terms of what people do well. There's the danger of it being too easily accessible in a certain sense, but then you've got this big process behind it. What do you going to do with public complaints? How are you going to deal with them?

MS. KIRKENDALL: That's another thing. We have our National Energy Information Center and they already talk to our customers. They try real hard to answer questions, to identify problems, to communicate back and forth with people. And that process has to continue. I think that goes part of the way to satisfying what they wanted with requirement three but then in addition they want this kind of record keeping on the complaints that we get. So it seems like we have to provide a formal way for people to complain. I want this complaint to be registered against this particular requirement. And that's why we came up with something like a button on the website.

A proposal that was made within EIA was to just ask our information specialists to write down when they thought it was a comment about quality. That's a nightmare. In the first place, everybody classified it differently and then we have to have somebody do some data entry. So who knows what you have if you did that. I don't know how valuable it would be. So I would think that having something on the web. If you would like to register a formal complaint or maybe different words, but some button on the web for this particular part of the process. I agree that we'll have to say something in *The Federal Register* notice about easy access to people for help and for discussing problems in general.

MR. SITTER: But I guess my question really was how do you propose to handle the button on the web? Somebody just is looking at the website or they've got some data. They do there. They put down a formal complaint. You have to now deal with that formal complaint.

MS. KIRKENDALL: We will have to deal with it. We'll have to have somebody assigned to look at each of those when it comes in and decide whether it needs a formal response. Depends on how many we get.

You probably need to make a formal response to everybody. It could be by email. Actually, you can't. If you do it on the web, they have to tell you how they are if they want a response. If you want a formal response, then tell us who you are.

CHAIRPERSON CRAWFORD: Did you have a comment?

MR. BREIDT: Yes, I would just like to echo Randy's concern that if you make it too easy, you could certainly suffer from that. It seems that EIA is fairly web-intensive so you have an easy button to allow people to complain, you get a stack of complaints that you deal with however you do it, but OMB sees these reports from the different agencies and you've got a huge stack of complaints. Are you going to be penalized because of that?

CHAIRPERSON CRAWFORD: Maybe you should have to make them call. I mean if they're really all fired up about complaining, maybe they need to make the extra effort.

MS. KIRKENDALL: Or write a letter. That's what I had originally thought was writing letters. However, in light of the current situation, I'm not sure.

MR. SITTER: It's an electronic communications problem. When people are upset -- I mean you get it sometimes. You get an email which is overly --it's just not what the person would have done if they were there or if they had to think about it for five minutes. It's a knee jerk reaction. You're not happy. You press a button. You press the button, it's gone. You can't get it back. You're not there in person so you can't say well, I didn't quite mean it that way. Blah-blah-blah. Maybe I'm wrong. I think there's a lot of companies out there with a lot of experience and you won't find too many of them that it's that easy to complain. Most of them require you to make a phone call or to fax them.

MS. KIRKENDALL: We already do that. It's not a formal complaint, mind you, but we give information about who to call, who to email. I think our people are really good about getting back to people, too.

MR. BERNSTEIN: I think you need to have

it on the web and I think that are ways to somewhat minimize. We've had enough experience with the web now and there's a lot of information out there on how to try to sort of not make it too hard for people to communicate but not make it too easy and to have sort of multiple checks along the way before they can actually send. There's not just a you write this and make a send button. You're going to have to go through a number of steps and then before you send it, it asks you to look at it again just to make sure this is what you really wanted. While it doesn't deal with all the issues, it begins to deal with some of them. I think you've got to give, in this day and age, you've got to give the opportunity for people to complain at the website. You just can't avoid that.

MR. BURTON: I agree with both these guys. I've often set up those emails you're talking about so I know how easy it is. On the other hand, particularly when I'm having a problem or there's some level of dissatisfaction when I'm dealing with a web-based activity and they say if you have problems, call us. Then I know what they're trying to do is avoid me and that just makes me that much more determined.

MR. CASSELBERRY: The OMB guidelines, when they talk about the complaint process, it's really for people to complain that the agency did not follow its guidelines. An example they used in their *Federal Register* notice is when the Weather Service puts out a forecast that it's going to be sunny and it rains. People may be unhappy and they may want to complain, but that's not really a problem with the quality of the forecast service if they had a model and a way to do it and that was their prediction. And so we will get comments and we might not report it all to OMB if people are complaining about things. If we put data out there and describe its limitations and whatever other aspects of it and we call it our quality guidelines and people just want to say that they don't think it's really a good representation of the real world, then that may, as long as we feel like we followed what we were supposed to be doing and we explained it, then we may not even classify that as a quality complaint. I mean not everything that's going to come in through this site will necessarily be reported to OMB.

MS. KIRKENDALL: I think the denominator has to be required. How many complaints did we get?

MR. CASSELBERRY: Yes, and then we'll say how many of them were valid complaints about our adherence to the quality guidelines and then something probably about how we resolved those, either just a discussion or something.

CHAIRPERSON CRAWFORD: Nicholas.

MR. HENGARTNER: Two things. First of all, in terms of complaints. What are the people going to complain about? The data? They're going to have to know a hell of a lot to be able to say that number is wrong. I wish they would tell you how they got to know that that number was wrong because that's what you're in the business of telling them.

The other possibility is the model is wrong. That already more subjective. If you have a good review and especially that's one of the things I might suggest you do is not only keep the information about the model but keep a log of all the changes that have been done on the model like every software has a log of the changes and when it was done by whom. Do the same thing for model because model is going to be updated. All your forecasting models are currently the best available and then something else comes up and you have to change. I mean September 11 is one example. So keep in mind those changes.

The last thing people probably are going to complain the most is the timely publication of the data, the procedure. I think if you can somehow distinguish between those three sources of complaints already on the webpage so that not all those submissions are going to come in through the same channel, that might help if it's simply this wasn't timely enough for us. That's good for you to know because that gives you an idea of how well you serve your public. If you said that number was wrong, that's a different type of complaint. I actually would like to see those things separated because it gives me an idea of how good I'm doing and to whom and so forth.

MS. KIRKENDALL: Right, and I think that if they complained about the data, it could be because there was a mistake in the data, in which case we're happy to know it and maybe we'll fix it, but we could have still followed our guidelines with all the reviews and the edits and all the things. So one of the things we're going to need to get from OMB is clarification about how you count that. We did follow our guidelines, and I think that means we don't have to count it against us.

MR. BERNSTEIN: But for getting particular information, if you get that situation, then you got to go ask why did the guidelines not catch it?

MS. KIRKENDALL: Yes. Well, that's true.

MR. BERNSTEIN: So if you get somebody that comes in and says I've found this problem with the data, how come that didn't get caught?

MS. KIRKENDALL: That's true. We would want to know it because we want to fix it. Same thing with models, too. One of the things that I guess it's Social Security says is they don't want to get criticisms that -- no, BLS, the CPI is wrong. I know that because my social security check is too low.

CHAIRPERSON CRAWFORD: Roy.

MR. WHITMORE: Randy was talking about some people might send in just very quick complaints without thinking about them and might want to retract them or in other cases, I think the opposite might happen. Somebody that's really thinking about something and might want to actually have somewhat of a dialogue. It might be useful when you've got the initial page where you're gathering the complaint information to assign a number to it for tracking purposes or whatever. You tell the person your complaint is complaint number so and so. If you need to write us regarding this, be sure to mention this complaint so you have a way of tagging together things and have somewhat of a dialogue associated with them. It may be that there'll be some cases where people would send in the complaint and then five minutes later figure out oh, well, this was the problem. I was wrong. If you got some information about how they were logged into the system, they could send back a retraction even or if it's a major problem, they can send in further information as they research the problem more.

MR. BURTON: That might also be a way to take into account what both Randy and Mark said. If their initial contact can be very easy through email but what your response to that says okay, we're going to give you a number now. If you want to go ahead and do something more formally, please write us and reference this. If all they want to do is blow off steam, they can do it quickly. On the other hand, if they want to do something more substantive, they have to engage in a real process.

CHAIRPERSON CRAWFORD: Thank you very much.

At this point, I'll ask for any questions or comments from our audience or anyone else from EIA.

MR. KNAUB: Can this be about anything? Fred Mayes particularly has talked on alternative transportation fuels. I was thinking. Does anybody know anything about network sampling because it sounded like the frame was a problem. Maybe you could tell us a little bit about how that might be appropriate. I remember hearing a couple of talks by Monroe Cirkin on that. It just sort of sounded like it might be something there.

MR. WHITMORE: Network sampling, snowball sampling, similar words. They refer to the same sort of thing where you're basically using one respondent as a means of identifying other potential respondents in a population. So certainly with regard to the converters to alternative fuel vehicles, one converter might well know about others that are in the same business, even if there's not a trade association. That might be a way of building up frame for those sort of things is using networking procedures.

Jay is nodding his head.

MR. BREIDT: Yes. -- along the same lines.

MR. KNAUB: The frame was sort of a problem. You don't have to know the frame ahead of time. Right? Is that correct?

MR. WHITMORE: I think in this case, the usual application of networking is within the sample. If you're doing networking within the sample, one sample member identifying additional sample members after you selected the sample, then you do have a multiplicity problem that you have to take care of in the analysis. Basically for every member of the sample, you then need to know how many linkages to the frame does that individual have. So if we're talking about saying one converter bringing in another converter because they know each other, each one is going to have to be able to tell you how many converters do you know. I don't think that's something you know very well.

If it's well-defined in terms of familial relationships is one way that that's sometimes done. You're looking for people that have a particular disease that has a genetic link. Then you ask well, how many within a certain relationship, aunts, uncles, cousins, whatever, have the same disease? And so you have a finite number of people that are kind of linked together and for the multiplicity adjustments, you can ask any one of them, well, how many other aunts, uncles and cousins, whatever, have this disease so you know the proper weighting factor to use.

For the alternative fuel vehicles, I'm not sure you'd be able to do that. You might be able to use some techniques like that in developing the frame. But in terms of implementing the survey, I think you've have a hard time coming up with those multiplicity factors.

CHAIRPERSON CRAWFORD: Any other comments or questions from the audience? Any questions or comments from the public? Before I adjourn the meeting, I just want to remind everyone of two things. First of all, our spring meeting is in March this year and you should already have the dates for that.

The second thing is that effective at the end of the year, I'm stepping down as chair of the committee and Jay Breidt, who's sitting over there next to Nancy Kirkendall, will be the new chair of this committee effective January 1. So if after this meeting you think that you have some suggestions for the next meeting for topics or presentations, you can contact either Jay or myself and give us those. We'd appreciate that.

Thank you. This meeting is adjourned.

(The meeting was adjourned at 11:17 a.m.)

CERTIFICATE

This is to certify that the foregoing transcript in the matter of: MEETING

Before: AMERICAN STATISTICAL ASSOCIATION COMMITTEE ON ENERGY STATISTICS

Date: FRIDAY, OCTOBER 26, 2001

Place: U.S. DEPARTMENT OF ENERGY
1000 INDEPENDENCE AVENUE, S.W.
WASHINGTON, D.C.

represents the full and complete proceedings of the aforementioned matter, as reported and reduced to typewriting.

Eric Hendrixson