USDA Approval of Beef Grading Augmentation Systems
Martin O'Connor

[1]Our next speaker is Marty O'Connor from USDA/AMS Standardization Branch to discuss USDA approval of beef grading augmentation systems. Marty has devoted twenty-six years of government service in the meat science area. In his present position as the Chief of the Standardization Branch, Marty is responsible for the development and revisions to the official U.S. grade standards for carcasses from red meat species. In addition to the meat grade standards, he is responsible for various meat purchase specifications and the development of USDA certified branded beef programs. Current major initiatives include instrument prediction for marbling and yield in beef carcass assessment, DNA technology for USDA/AMS marketing programs and export verification programs for major international beef markets. Let's welcome Marty O'Connor. Thanks Deb. One thing I need to clarify. I did not pull the plug on Glen's PowerPoint presentation because he was going to put up the grading statistics on what I did when I was grading. So I just want to make sure that we're clear on that and Glen stay away from my presentation.

[2]Probably cover a few things that have already been discussed by others this morning. Both Glen and Eric talked about some of the issues related to technology and the measurement of it. But one thing that Glen talked about and I think is worth reinforcing is what we're looking to do in the approval of these instruments is to reduce the within plant variation or the variation between plants. That's one thing that he pointed out in some of the data that he presented. They have discussions ongoing with us all the time about that. That there's geographic differences or in-plant differences. Because in some locations in the Dodge City area for example, we have fourteen graders located in that area. So we have fourteen different subjective evaluators that come in and go out of that plant. And like he says, they had approximately fifty USDA graders in all their plants across the nation doing evaluations for both quality and yield in that aspect. So we all recognize and understand that there is variation due to the subjective nature of the way that we are applying these grade standards at line speeds that are a little bit different than Len was doing there in his movie where they weren't even looking at the ribeye, my goodness. And so as chain speeds have improved or increased, excuse me, the decision-making process of the subjective evaluator has also been challenged. There are more things that we are doing online. There's different certification programs. There's a variety, we have over thirty-five different certified programs that we're currently providing certification for online. One of the things that we've done to mitigate the errors in online application at these high speeds with multiple specifications is having two graders on the line, conducting these activities. So one is either assigned to marbling and/or certifying for branded programs or the other's doing yield and a combination, whatever fits best for the unique situation that is apparent at that location. So we've tried to modify the best that we can in order to address the concerns of the high-speed accuracy that needs to be implemented in a subjective manner. But we are at the twenty-first century and we're definitely want to embrace the movement toward technology to have more of an objective measurement of the integral parts of the determinations that we're about to make. Another point that Glen brought up in moving up and down the supply chain or providing information or paying for cattle is the ability to be more precise in our determination. And in that we can segment where the true value is associated both in quality and yield by segmenting the marbling degrees and/or the yield grades to a more defined true value segmentation to a tenth of a yield grade or ten degree marbling units, whatever the call can be. And that can, once these instruments are approved and in place, can in fact be variable between plants or for different specifications. We see this as augmenting not only the grades but also the certification activity for branded programs that have grade characteristics as part of their program too. So there's going to be a lot more ability to segment into more heterogeneous groups into those smaller homogenous settings that will allow that to be segmented out either monetarily or for brand specific specifications.

[3]And like Glen pointed out too I'm sure he's talking about me in the top left where I was both accurate and precise and not, and not the one that was down in the

lower right-hand corner. In our subjective determinations and oversight of the grading activity, we like to strive for the top left. We hope if we're not there, maybe the top right is our next category because if we are precise, it helps us to bring those people back into accuracy easier. If they are precise, it's easier to move them then if they're all over the board. So if they are precise, at least we know how to move them. But if we take our own measurements and see they're all over the board, it's a little bit better to get them both precise and accurate. So again, we strive for the accuracy and precision and recognize that we're not always there but hope, when we do make our corrections, the correction is done in the manner that'll bring it back into that arena.

[4]A little bit of a history of the grading instrument technology. It really first started in 1970's where the first funding was provided by the Livestock and Meat Industry Council. Kansas State University was doing the work and Dr. Dell Allen, who's in the audience this morning was the initial investigator in that. Part of that research was done with information and technology from the Jet Propulsion Laboratory. I think the lead engineer. He was telling me yesterday from, electronic engineer from Kansas State had contacts and had previously worked at the Jet Propulsion Laboratory. So that was kind of the connection on how that was, got the connection there and got moving in that direction and movement. Then in the 80's, Dr. Cross at, when he was at Meat Animal Research Center started to do some testing of the prototype that was developed at Kansas State. And he did the ninth, tenth, and eleventh twelfth rib dissections for prediction of yield. And all this work here is talking on yield prediction. And after that, at Kansas State done in Wichita later on in the eighties, whole carcass cutout determinations were evaluated for determining the precision that those instruments could bring to the industry.

[5]In 1994, NCBA coordinated the National Beef Instrument Assessment Symposium led by Dr. Reagan. This again reemphasized or brought back to the forefront the issue of instrumentation and the need to move forward in an objective manner in measuring the important characteristics of quality and yield in the grading of beef carcasses.

[6]Finally we got into a new millennium. We were still in the subjective mode of evaluation. As Eric says, the more we go along, the more we stay the same. We've always embraced any new technologies or any new approach in doing the betterment of any assessment that is, that we do the provision of in the area where it is a voluntary service. We recognize there has to be value associated with any services that we provide. Otherwise there will be other technologies or other ways of attaining what needs to be done that is more cost effective or cost beneficial to the requestors of our service. So we want to be out on the forefront of any technology-driven requests or research to insure that it can be adapted to what we're doing or that we can adapt to use it to supply a meaningful approach to prediction. So with the new millennium in 2003, we published a standard for the use of the instrument, video image analysis instruments to predict the ribeye size. Once that was approved, we had two systems, the RMS and the E+V system approved under that. It allowed those to present both carcass image or excuse me those ribeye images or the measurement of those ribeye images to us in determining our yield grade and the official yield grade evaluation. In research that was conducted in the nineties it was determined and as everyone probably knows that ribeye measurement or ribeye estimate is the most variable factor in the yield grade equation especially when you're at five hundred carcasses an hour coming down the line. The graders, even though they're doing a day in and day out and do have a pretty good judgment on that with the variation of the cattle coming through there that is the factor that had the hardest to attain consistency, precision and accuracy. And so that was the first issue that we worked on. We felt it was one that could be tackled and maintained and brought forward first. So in 2003, we published that and I believe in 2003 and 2004 both instrument providers had passed the testing for that. Last year we had finally developed a performance standard for the complete prediction of the yield grade. And have tested two systems to that manner too. And have, they've both passed the functionality portion of being able to apply a yield grade to the USDA grade, yield grade standards. There's another segment that they'll need to address and I'll talk about that in a few minutes. It ties into what Len's discussions earlier on ASTM. But it's the, what we call

operational procedures or how you bring forward the ability to operate that camera on a day in, day out basis and in the manner that it was approved in the research setting or the testing setting. And then just this month, we have published a performance standard for marbling. It's as everyone probably knows, it was the most hard area for us to get a grip on and determine what is the best performance standard as well as the instrument providers on how they can accurately and precisely predict a, the marbling since it is not always and it wasn't an objective measurement at all of what a grader does in that assessment. There's three main characteristics in developing a marbling score. And with the objective measurement it's probably easiest just to measure quantity but there's others, distribution and texture of marbling too that's taken into consideration that must be addressed. And when emphasis is, was put on developing an approach on how to account for all three of those variables in determining a marbling score, I think again the research scientist in this room, a lot of them took this as a challenge. Went forward in trying to define what we can do in the predication of that marbling score and get moving and working with the instrument providers. Even just last fall at the AMI Food Show in Chicago I was talking with one of the instrument providers. And he was kind of asking me what we felt the performance standard would be and I told him, you know, I was expecting an R2 of point eight five or higher. And he said, oh man there's no way that we're going to be able to meet that. And I said well, you know, that's kind of what we we're doing on the yield side and hopefully, you know, we haven't finalized that yet but that's what we're identifying to go forward on a quality side. And so he said, you know, they went back from that point on and worked extremely hard and I'm getting a little bit ahead of myself to the point of where when we did actually publish a standard for them to perform against that he was prepared and ready to go after, but it just took an initiative and fine tuning where they were going to go with it to get to it.

[7]In the development of a standard there's, insurance is especially from government perspective that we get input from all segments or all sectors to insure that we have a wide variety of information available to us when the ultimate decision has to be made of what we expect the final performance standard to be made. And in order to do that, we've developed a task force with the industry that included both USDA, National Cattlemen's Beef Association, the manufacturers of the equipment, academia, as well as the beef processing companies. In each of the processes, whether it was ribeye size, yield grade or marbling standard development, this task force provided us invaluable information, refinements, comments in order to insure that in fact we had at least all the information available to us to make a sound decision. Everybody might not have agreed what we did ultimately in each of those standards. But at least input was gained from all segments to insure that we had the sound basis to make the decision on. And then as everyone says, well let the government do what they want to do. But we did receive very good comment and it had a very good working relationship with all the people in each of those segments.

[8]Some of those that were instrumental in providing the input or working on these projects are listed here. Might want to just point out a few people that are in the audience today. Jimmy Wise, who is now retired from USDA was working hard early on to get this thing off the ground, both on the ribeye size and the yield grade program. Kari Smith who's presently on staff now has done a lot of work on the marbling component. Chancy Moore is also in the audience I believe and Bucky Gwartney in the back from NCBA as well as the technology providers as well as the companies and we're going to get into a couple of trials. I especially want to acknowledge Swift out of Greeley and Cargill out of Dodge City that provided a lot of assistance to us to insure that as we progressed in this area that they made space and carcasses available to us.

[9]Not only did we use internally or domestic input and analysis for determination of our standards development, the European Union has had an approved program for vision system classification in effect for several years and especially in Ireland since 2003. We also talked with the Federal Research Center for Nutrition and Food in Germany, the Ashtown Food Research Center in Dublin as well as the Livestock, Beef and Sheepmeat Division, Irish Department of Agriculture and Food in Ireland, where they have twenty-four vision-based systems in place and approved to do carcass

evaluation. It's a voluntary program in Ireland but it's pretty much compulsory by the requirements of most of the producers will not take their beef to any place that does not have a vision-based system. They just completed an EU evaluation of the instruments after they've been implemented since the, actually 2004 when they first put the first implementation of an instrument in at that point. Then the European Union came in with their first audit, complete audit of the system. It was in June, no excuse me it was in May of this year. And the results just came back a couple weeks ago that they did pass and met all the requirements of the EU standards for vision-based instrument as well as predictive accuracy.

[10]As I discussed earlier, especially in the marbling component and that's what I'll probably spend most of the rest of my talk on is talking about the marbling inspection or prediction of marbling is, right now a very subjective manner. And how we can train instruments or vision systems to replicate what a person sees visually has been quite a challenge. There's some, was some discussions in our task force groups. Well should we just take extractable fat and set a standard against that and measure how that would work. And use that as the benchmark for the determination of the acceptability of his instruments. And as discussed earlier, just amount is not the only thing that is considered once we're looking into the grading. We need to do the distribution as well as the texture.

[11]Again working with the task force, we developed some studies to establish the performance standards that would or finally become published as the official standard. And we enlisted the assistance of both the main technology hardware providers to assist us in this endeavor both E+V system out of Germany and the RMS system out of Fort Collins, Colorado.

[12]We published in January of this year a notice to the trade that was accompanied by a draft standard and protocol on evaluation of instrument systems. What the blue document that overlays the notice to the trade there is our first official grade standards that were published in 1926. Orion Samuelson was talking about change and how we need to be prepared for it and do things. And he was analyzing a hundred years back or a hundred years forward. Our grade standards have had some changes. We've had thirteen changes in the last eighty years in the grade standards. Some of them significantly, probably a couple of the big ones when we ribbed carcasses and started looking at the twelfth, thirteenth interface of the ribeye at that point and the other, the introduction of yield grading in 1965. But I think those are all portions of the standard itself, the application of the grade standards has been pretty much the same from the beginning of the time that we've actually provided them in the voluntary state all the way through the present. But what we're looking here is a fundamental change in the way that we do the evaluation. And we believe as Glen had pointed out earlier today too and as everybody probably recognizes an objective measurement for consistency, accuracy, precision, uniformity, whether it's perceived or true, is going to be a big assistance to the industry to take us forward to the level of competence that we need to have in order to maintain the integrity of the grading system as we have it presently.

[13]As I said it was presented to the industry for comment in January of this year. We had held two or three different conference calls in order to obtain feedback prior to conducting two separate trials to insure that we are collecting, we're collecting the right information or looking at the, evaluating the information in the appropriate manner or had the design of the trials set up appropriately. The performance requirements, we changed a little bit from what we had done in the past for ribeye size and yield grade. And we went to the, based on regression analysis instead of more of a comparative analysis. We thought this would give us a better evaluation and a better way of approaching and approving these systems. And then we conducted a few trials in January and February of this year. And again those two trials conducted, one's in Greeley and one in Dodge City, Cargill.

[14]There's three components of the draft standard that we had developed. The first was the demonstration of stationary repeatability. How those instruments can do on the same carcass in a repeated manner, a three-time repeated manner. Phase 2 is demonstration of accuracy and precision at line speed or operational procedures.

We're going to test them actually at line speeds. And then third is our operational procedures or the performance standards on recognizing the ASTM standards as part of the performance that people need to be trained, calibration has to be done. How the instrument was installed. And these other issues that would impact the ability of those instruments to take an image and analyze it properly and correctly.

[15]This, our first standard, again based on regression, we had an R2 of point nine. So the instrument provider that I told point eight five, he was even more worried when I, we came out with the final of a point nine. But even he'd been in on all the conference calls. So he also called me after each one of them. The, then I don't we went with the linear approach in determining precision and accuracy and also the root mean square error of thirty-five marbling score units.

[16]Some of the protocols in the trial. Every carcass for the whole day had to be imaged by both instruments. So it was a real life situation. They weren't just cherry picking the carcasses that we would identify for the test. Every carcass had to be imaged and after each provider imaged the carcass, then we'd make our selection as to which ones we're going to include in the instrument evaluation. And through, as about a two and a half day process, independent officials, grade data was collected on two, a little over two thousand carcasses. And the expert panel was made up of three USDA people. And then data was analyzed at a multiple different ways. Once we'd collected it, again this was our first foray into marbling. And so we wanted to insure that the draft standard that we had was either adequate or it demonstrated what we needed to look at in order to refine it to make it more useful document and something that would withstand the scrutiny of all segments of the industry. Here we're venturing into a new arena. And again marbling's probably the biggest one on the block that everybody's really wanting to look at and we're moving very judiciously slow in this that we want to make sure that we move, bring something to the forefront that the industry can use that is embraced by all different segments and it is accurate and is precise as the subjective evaluations that our graders are currently doing. And we don't want to do anything that would set us back ten or fifteen years if we came out with something that might have some bias to it or might be perceived as not being as accurate or precise as the subjective evaluation or having a big hole in it. So this is, was one of the biggest concerns we had in going into these first two trials. Some of the discussions that we had in our working group was that when we did develop a standard or a draft standard that it had to be something realistically that we had to be able to measure that against what the line grader is currently doing or how they are performing now. And not have an unrealistic expectation of an instrument to predict something that it is being designed with input from the subjective evaluators to begin with. In 2003 I think we worked with RMS in Cactus, Texas, just as an initiation I talked with the chairman, Bob Richmond, just saying we were wanting to move forward more and more into the area of instrument evaluation of marbling. Even though at the time we were working on yield, but we wanted to get some data just and analyze at that point. So they agreed to work with us and Swift volunteered their plant there so that we could do some research there. And we collected data for I think it was a three-week period. Just it wasn't a real set design on how we were going to do that. But we rotated in different supervisors from across the country and made dual comparison marbling calls at that location. And then a year later in talking with Dr. Keith Milk (SP) at Colorado State. He had a person that was very interested in looking at the data from that cause we hadn't done too much with it at the time. We were a little bit short staffed and stuff. So Chancy Moore at that point started looking at the data and felt that there was a way that they could develop a equation that could accurately predict the marbling score in association with a subjective evaluation that we currently were doing. So that kind of initiated that next round of push to get the standard published or the draft standard published. And then we were in communication with the other provider, E+V and they too showed the interest but had some concerns that we might not be approaching it in a most judicious manner. But through the working group working with NCBA and us in that working group and all the other segments of the industry, we designed this performance standard and moved forward at that point.

[17]But at the conclusion of those two trials, we recognize that both the expert

and the instrument marbling scores were to determine with some degree of error and that the variation with the instrument and the expert increased as the marbling score increased, which you would probably expect again the subjective evaluation and the instrument being trained against those calls that are made. They kind of moved in parallel as the marbling score increased.

[18]So after analyzing the data in several ways and working with a USDA statisticians as well as several researchers on that committee, we decided to abandon the regression analysis based upon several of the significant tests that were performed by the statistician and evaluations and went back to more of a comparative analysis.

[19]We weren't quite sure what the most appropriate approach was in doing a comparative analysis we, you know, had history with the yield grade and the ribeye size prediction. But we had run across a paper from Altman and Bland that kind of fit the way that we were trying to look at testing and instrument and its ability to predict the marbling score that was evaluated by us. So we based the development of our final standard on this Altman/Bland approach.

[20]After we worked with that group and developed it, we published it again in this month and some of the basic performance standards of that are that number one for the demonstration of the repeatability, accuracy, and precision and in phase, what we all the other portion is the implementation and verification of operational procedures. Again that's the ASTM or how the people are trained and how things are implemented. So number one is really the functionality of the hardware and the software. And number two is the company's approach in implementing that technology and once it's approved.

[21]Changed a little bit for the final testing too. We determined through evaluation of the data that we needed to use multiple plants and came to the conclusion that four would probably appropriately represent the differences that we needed to have explained in any evaluation of the data. It's still going to be, we're going to collect one thousand carcasses from each of those plants so we'll have four thousand, a minimum four thousand data points to evaluate upon the conclusion. We're going to split that data set or that data into two different sets. One will be a calibration data set that we will provide the instrument providers so that they can develop an equation. And we will test those equations against the sequestered test data set. That'll be the other half of the data. And what we're going to do on that is on a daily basis of evaluating marbling in each of the locations, we're going to rank order all the marbling scores from lowest to highest and start at the top. Throw one into calibration the next in the test and do that each day of each plant locations to insure that we take into account all the effects that may take account for on a daily basis. We also increased the expert panel membership to five. Three of us was too easy to beat up on each other. We felt that five would be a little bit more fun. But in actuality, we think we're going to get a more precise or a truer value for what the expert marbling score would be in that approach.

[22]So we have the standard that is published out there. Last week we conducted the first segment of the four trials to do the testing of the equipment. It was conducted in Greeley, Colorado.

[23]And as I'd said, there are two segments of the functionality. One is the repeatability and one's the online segmentation. The repeatability, ninety-five percent of the carcasses shall have at least, shall have all three of the predicted marbling score observations within twenty marbling units of the mean predicted of that marbling unit instrument. We've concluded that both of those instruments have passed that Phase 1 approach. So they don't have to do that at every location. That just happened to be demonstrated on one set of carcasses and we completed that. And both of them have maintained that or met the requirement for the repeatability phase of the approval process. But then the hard part's ahead of them now I think.

[24]The online or functionality requirements have average residual of zero plus or

minus ten marbling units. Standard deviation of the residuals of less than or equal to thirty-five marbling unit scores and have a slope of zero with the dependent variables, the residuals from instrument and the expert panel. An independent variable is the average of the instrument and the expert panel. Again we've run models of this standard as well as the other standard against correlation data that we have had from supervisor meetings as well as individual graders that have been performing. And we have those that do meet that individually and as a group as well. So we feel that the standard is attainable. It is approachable. And the instrument providers feel that they are confident that they can meet performance requirements that are set.

[25]So now we need to implement this technology. This is a picture just of a trial that we had conducted earlier where we both had the grader looking at the carcasses and in him inputting or augmenting his grading activity. But as pointed out earlier by Glen--

[26]There are some things that camera vision system cannot do. And one of them is evaluation of maturity right now or dark cutting calluses--

[27]Or color, texture or firmness, even though we recognized color can be evaluated. It's not looked at in the approval process at the present time as something that can be looked at from the USDA perspective as given a maturity score in that segment.

[28]Then we moved to what we call a prime 2 or again the requirements on a plant to perform once the instrument is approved.

[29]These are on a plant-by-plant basis. We do reference an ASTM standard, which is a procedure for use of the requirements of the last meat and poultry evaluation devices or systems. Whereas any good government bureaucracy, we also have additional requirements. But I think in fairness to us, Len did point that out that the ASTM standards are just the basis for development and you, he would and anyone would expect to have other user requirements built into any system for approval.

[30]Some of the issues like Len had pointed out already for, that are covered under ASTM are installation, training, calibration, testing, maintenance, traceability. It isn't expressly written into the standard however, we have reference to it and it's kind of globally expressed in the ASTM standard. But it will be an extreme component in the approval process that we give credence to.

[31]The additional AMS requirements are additional operational requirements of some specific things that we'll need from the plant to demonstrate their competence in providing carcasses that meet the requirement of the standards for both for evaluation as well as optimizing that information that can be obtained from them. We need to have cyber security addressed. We have to insure the integrity of these systems as we move to a more technology-advanced way of evaluation to insure that the equations that are approved for presentation to us are secure. That we have a copy of those and that we benchmark and check against the raw input data that we can come up with the same outputs as the instrument to insure that in fact the integrity is there. That only certain people have access rights to make any specific changes to the instrument itself. The revision systems as well as making sure that the equation and algorithms are locked down. And of course the manufacturers also have a hand in that as they want to make sure that their proprietary information is in secure form. But that's something that we'll be critically looking at because we also based a lot of our reports and information off of the outputs of the grades that are being performed. And if we're going to rely upon these instruments to make those evaluations as well as report them, we have to have the integrity of the whole system in tact. And another key component again is a little bit more along the same lines of cyber security is identification of the hardware and software components. We will have them, they'll have to have a listing with us and inform us of any modifications, changes, or things that they do that would impact the integrity of the system or the capture of the information.

[32]One of the additional requirements that I wanted to give you an example of would be ribbing. And that would be the proper procedure and making sure that the saw cut is proper, the knife cut across the twelfth/thirteenth interface of the ribeye is appropriate as well as the time I think. I forget. It might have been Eric that was talking to them about bloom time in making the evaluations. And that's a critical component that we see in these evaluations, even though we have a minimum time in our standards that what must be met for someone before a subjective evaluation is conducted. We're going to insure that there's a uniform and an optimized amount of time that is done for presenting these carcasses for image capture with these video image analysis.

[33]Future opportunities. Again, like Eric said, we don't want to be stagnant. Orion Samuelson said, be ready to change. We need to change, embrace change. We think there's continual improvement for the services we provide. Again we're a user fee or voluntary service. If we become antiquated in the approach that we take or the services that we provided, there will be no need for us. And that may be good or it may be bad, but I think for that independent third party evaluation by the government, there's still some need for that. And people still embrace it. And we want to be prepared to provide the services that are requested of us. Other technologies for tenderness, is it near infrared, genetic markers, or whatever. Are there are like some folks were saying earlier today, are we even looking in the right arena? Do we need to expand what we're looking for and doing and make sure that in fact that we will approach palatability in the, in a manner that will express the best method and giving the highest reliance on any approach that we do take.

[34]Just show you a quick image of both technologies here. On the right is the actual instrument itself being placed over the cut surface of the RMS technology. And on the left is just a, the video image enhanced vision of what some of the measurements are that they are taking.

[35]In addition to our cold system, they also have a, what we call a hot system. And again that's the European model is a hot vision system that is being conducted over there for their carcass classification. They use confirmation and they use fat test classification. But the same principles in oversight pertain and that's, really was what I was there looking at is their ability to maintain the integrity of the system, checking it and calibrating and validation and not necessarily the prediction of their classification system.

[36]Here's the E+V system. A little bit different design in the technology itself but both functional in the, in the same manner. And then there's an output image on the left.

[37]They're hot, they have a hot carcass system as well and again that's one that's based upon the European Union standards and predictive modeling over there.

[38]And I'm off center a little bit there, but that was supposed to be a video of what we're currently doing now. How we are subjectively evaluating and applying the grades. And then we transition into the new phase of technology. It's a little hard to see there. We're off center for some reason. But he's got that counterweighted instrument that he's placing down over the twelfth, thirteenth interface. And you can see in the background there that the images are coming up on the, a screen that's right next to him so he can insure that the quality control checks are in fact accurate and he could retake the other side if need be. You'll notice he's doing just like Glen pointed out, they're just taking a one side at the present time to make the classification. This is the E+V system. Again a little bit different manufacturer of the hardware itself, but working on the same principle. But again the predictive modeling is what is different. This will painfully be over in a second here, unless you can.

[39]I think my last slide was just giving the website where we have our standards posted.