F-D-C REPORTS —

FOUNDED 1939
$595 A YEAR

"The Gold Sheet"®

PHARMACEUTICAL & BIOTECHNOLOGY QUALITY CONTROL

THE NEWS THIS ISSUE

Vol. 39, No. 10 October 2005

STABILITY PROGRAMS FOR BIOTECHNOLOGY PRODUCTS may require a reassessment of development and early clinical materials to understand the significance of process, product and analytical changes made later, FDA and industry experts are cautioning. At a "CMC strategy forum" on biotech product stability, participants explored the intimate web of relationships between the analytical work done in the development phase and a viable stability program on the commercial product. Producing and retaining enough material during the earlier phases is among the ways firms must plan ahead if the program objectives are going to be achieved. Challenging issues for biotech products include: • how to set reasonable stability targets in the early clinical phases • the role of stability assessments in change control • the implications of non-conforming results, and • how to incorporate new information as assays improve. *[Included are presentations from the stability strategy forum on the role of stress studies in the development process* (pp. 4-6)*, setting up a stability program in the late clinical/application phase* (pp. 10-13 and 16-20)*, and post-marketing issues* (pp. 23-28)*].*

FDA And Industry Wrestle With Vagaries Of Biotech Stability

The potential importance of retaining sample material from development and early-stage clinical batch production for later evaluation is being emphasized by FDA and industry experts in exploring how to improve stability programs for biopharmaceutical products.

The value of being able to reassess early production material as processes, formulations and particularly analytical procedures evolve was a recurrent theme at a two-day "CMC strategy forum" on the "design and successful implementation of a stability program for biotechnology-derived products," held at the National Institutes of Health (NIH) in late July.

The stability meeting was part of an ongoing series of CMC strategy forums for well-characterized biological products (WCBP) being coordinated on a semiannual basis through the California Separation Science Society (CaSSS). The forums are designed to bring industry and regulators together to discuss ideas and share experiences on key issues of concern, with the goal of developing better CMC-related standards and guidance in the evolving biotech arena.

Parallel workshops focusing on viral vaccines and biotech comparability issues, respectively, were held in January in conjunction with the annual FDA/CaSSS WCBP conference in Washington D.C. At NIH the previous July, the topic was process and product impurities ("The Gold Sheet" September 2004), with the January '04 forum addressing specification setting ("The Gold Sheet" February 2004). The forum coordinating committee is publishing white papers on each of the topics based on the presentations and extensive discussions at the forums.

In coordination with the annual CaSSS WCBP conference in January 2006 in San Francisco, a CMC strategy forum will be held on the use of reference standards to support product and method specifications for biopharmaceutical products. The forum at NIH in July '06 will focus on the "changing paradigms in process validation."

As for the other forums, a series of questions was developed for the stability meeting to focus the discussions and help shape a more meaningful consensus position.

➢ The discussions and presentations which preceded them on the challenges of designing and implementing a stability program were organized into the three stages of the product life-cycle: ● early clinical ● late clinical and marketing application, and ● post-marketing.

Concerns receiving particular attention for early clinical development included the basis for selecting methods and for determining expiration/retest dates. Issues generating debate for the late-stage clinical process included the handling of small-scale studies in evaluating drug substance stability, bridging study considerations when assessing manufacturing and analytical changes, and regional differences in stability

testing and filing requirements. The post-marketing discussions centered on defining and handling non-conformities in stability studies and the role of stability in change control.

➢ Through the course of the stability forum, participants explored the intricate web of relationships between the analytical work done during the development/early-stage clinical phases and the establishment of a stability program that will be viable through the lifecycle of the biotech product.

Producing and retaining enough product during the earlier phases was recognized as an important component in assessing the linkages and helping to resolve the issues that arise later as new information becomes available and changes occur.

In summarizing the CMC strategy forum discussions at an IIR-sponsored workshop on "stability-indicating and forced degradation assays for proteins" held in conjunction with the forum later the same week, Biologics Consulting Group Senior Consultant Nadine Ritter noted the emphasis placed on the retained samples at the forum. Ritter serves on the FDA/industry planning committee for the strategy forums and will help in preparing the white paper that will be developed for publication based on the stability forum discussions.

➢ The message, she said, is to save the earlier material "even though you don't know the real-time stability of it. We can set stuff aside in development and early clinical trials and then come back and test it."

Ritter recognized that "there may be some artifacts that are the result of degradation at the state that you held it." However, she said, "those can be adjudicated with a stress study and be proven that they are actually the result of some intermediate degradation that occurred." Pointing out that "assays do change over time" and "your measuring stick is variable early in development," Ritter stressed that "by retaining some of those materials, you have the ability to go back and look at them with tighter methods if that is what you want to correlate to."

Another participant at the workshop commented that preserving samples that had been placed on stability, particularly in a real-time protocol, can also be valuable.

"If you can demonstrate that your product is stable at, say, -70 or -20 ^{o}," he said, "then you might want to store those samples for retains if you want to assess whether any future methods might pick up different impurities and so forth, so you have actually some additional real time data – not just the retains for clinical trial materials or development tox studies, but also maybe even [from] stability studies, whether accelerated or real time or frozen."

➢ At the CMC forum, this sample-freezing approach was also recommended as a viable one.

"Early on, we don't have enough experience with the methods. The methods are changing and you get a lot of scattered data," commented Rinat Neuroscience scientist Roberto Rodriquez. "One practice that we have in development," he said, "is to freeze all the samples at different time points, retain the samples, and then analyze them all at once with the methods that we think are stability indicating."

Human Genome Sciences Pharmaceutical Sciences Group Executive Director Thomas Spitznagel concurred that the approach of freezing retained samples has proved valuable for his company as well, particularly during the development stage. "Even in the GMP studies, we like to have a little left over, freeze it down, and where issues come up, where it is hard to

"The Gold Sheet"

FOUNDED 1967
F-D-C REPORTS, INC.
5550 FRIENDSHIP BLVD., SUITE ONE
CHEVY CHASE, MD 20815-7278
PHONE 1-800-332-2181 FAX 301/664-7258
AN ELSEVIER COMPANY

Executive Editor: **Bill Paulson**
Composition: **Ting Yang**
Production: **Phil Simms**
Advertising: **Nicole Deschamps**
Customer Service: **Emily Brainard**
VP-Sales & Marketing: **Michael Magoulias**
Director of Web Publishing: **Patrick Bobst**
Editorial Director: **Melissa Carlson**
Executive Director of Operations: **Jim Chicca**
President: **Mike Squires**

"THE GOLD SHEET" (ISSN 1530-6194) is published monthly by F-D-C Reports, Inc., 5550 Friendship Blvd., Suite One, Chevy Chase, MD 20815-7278. The annual subscription rate is (1) $595 or (2) $345 for additional copies mailed in the same envelope with $595 subscription. Periodicals Postage paid at Bethesda, MD and at additional mailing offices. POSTMASTER: Send address changes to "The Gold Sheet", 5550 Friendship Blvd., Suite One, Chevy Chase, MD 20815-7278. **In Japan, Elsevier Science KK** is the exclusive subscription representative for "The Gold Sheet." E-mail: esm@esmixjapan.com. URL: www.esmixjapan.co.jp. Telephone: 03-3589-5290. Fax: 03-3589-6364.

interpret what a trend is, there have been cases where we have gone back…on the development side to take a look at" the question of whether it is "a real trend versus just noise in the data. And it has been useful."

During his presentation on "strategies for developing a comprehensive stability program for early-stage clinical products," Spitznagel pointed to improved sample handling as one of the main benefits of doing freeze/thaw stress studies (*see box, pp. 4-6*). "This is not necessarily for your stability-indicating nature aspect, but it is really going to help you with sample handling," he said.

Human Genome Sciences likes to do a variety of conditions and temperatures and to track by sizing as well as SDS-PAGE. The testing is informative if the bulk is going to be frozen. "But most importantly," Spitznagel emphasized, "this allows you to freeze your samples, [which] can really help dampen the noise of your assay variability."

Sample handling, he continued, "ends up being very critical in early development." In turn, "being able to freeze samples gives you some flexibility in the inevitable assay downtime. And finally, it also allows you to have meaningful retains. Early on your assays may not be fully developed. It is nice to have samples left over from your earlier stability studies when a new assay comes up. You can actually generate several years of data if you have freeze/thaw availability."

➤ At the follow-up workshop, Lilly R&D advisor Jerry Lewis commented that smaller biotech firms, which often have less material to work with and less clinical delivery flexibility, may have difficulty in developing the more extended sample retention and evaluation procedures.

The smaller firms "usually have just enough to do their clinical trial," Lewis pointed out. Under trial exigencies, the material "may be shipped to the clinical site before it is officially released [from] quarantine, and it goes into the patients immediately. You may have just enough to put into your patients, to have a little bit left aside to do some formulation studies, and the process development folks are kind of left hanging because they don't have any material to play with at all." Noting that this scenario is "what most companies are facing," Lewis suggested that it is not one in which there are retains readily available or sufficient flexibility to vary the age of material going into clinical trials.

➤ Ritter responded that "fighting over the amount of material that can be made early in development or who gets it" is also typical at the big biotech companies. Having the amount needed for the CMC evaluation, she stressed, is a critical issue.

"Certainly it is always an issue to be able to have enough material" to not only put in patients, but also "to get the information you need to make the right decision about the product," Ritter said. "Whenever I get in these arguments myself with project teams, I tell them it is not useful if all you can do is get enough of it to be able to get the clinical data, because without the CMC data, without the stability, it is not useful, and in fact it is possibly even dangerous because you don't know enough about it. So you have to fight those fights."

In terms of retains, Ritter added, "one of the things that stability people sometimes do that makes them either heroes or villains, depending on what side of the equation you are on, is that when they build their stability allocation, they put in overages to…account for mistakes." In turn, "sometimes these overage amounts at the end of the program become pots of gold there where people can come and get their retains if they haven't done it any place else."

Improved Assays Create Problems and Solutions

At the CMC strategy forum, the value of retained samples from earlier development and clinical phases was noted at several junctures during the extended discussions that took place on how to interpret and compare new information on degradation pathways flowing from process, product and analytical method changes.

➤ The point was made, for example, in the context of the more general concern of assessing the value of new or improved assays and the additional information and issues about the product that result from their application.

Addressing this concern further at the follow-up IIR workshop, Lilly's Lewis commented that new and improved assays developed after the clinical trails have progressed may reveal changes in the impurity profile. The question is then raised whether these impurities "existed all along" or "occurred during storage. Was it in the clinical samples and what relevance does it have because your clinical trials came out okay? So you end up with a conundrum:...If you create new analytical assays and start looking for stuff, you are going to find it. But if you find it, is it clinically relevant?"

Industry Experts And Guidance On The Role Of Stress Studies

The importance of stress studies in the development of a stable product and successful stability program was highlighted by several participants at the CMC strategy forum in July. Human Genome Sciences Pharmaceutical Sciences Group Exec. Dir. Thomas Spitznagel and Genentech Early Stage Formulation Scientist Mary Cromwell provided the following discussion of their role and the related guidance as part of their presentations at the forum.

SPITZNAGEL: Stress studies are probably the most critical aspect of a good development program. Why do we do stress studies? It really is a critical component of formulation development because it helps you identify how susceptible your product is to degradation. It also helps set sample handling and shipping needs for a stability program, and it certainly provides evidence of [the] stability-indicating nature of analytical methods. That is probably the most key point and that will come into play when we do our assay qualification.

Let's go through some examples of the stress studies that we typically perform early in development. The first one is a **shaking study**. We simply take our finished drug product in a liquid state, place it on its side, and shake it as fast as you can – typically over several time points, typically at room temperature.

This is an SEC chromatogram that shows a variety of time points. You can see the aggregate peak goes up slowly but steadily. In a very short period of time, you have shown two things: Number one, your product can aggregate – that is certainly a good thing to know; but number two, you know your method is capable of resolving those two species and that this is likely to be a good stability-indicating method. Sizing is typically the only thing we do with shaking, although there are other assays that can be used. And in addition to the assay, you can also find issues about fill/finish, shipping and mixing. It can also help you screen excipients such as surfactants.

Another stress study we like to perform is **oxidation**. Here we use a peroxide. We have also used t-butyl peroxide for surface exposed residues – typically around .03%, although this will be specific to your product of choice. We look over about 24 hours. Here we normally run reverse-phase primarily, although ion exchange can also pick up oxidation from time to time. Like sizing, it helps you identify the region of the chromatogram that is going to change during stability. These studies don't actually tell you what the peak is. It is a probability perspective here.

This is actually a reduced monoclonal antibody reverse-phase method. You can see the heavy chain here. Overexposure to peroxide forms a singly, doubly, triply (you can not really see the quadruply) oxidized species. Once again you don't know that without the MS data attached to it. But just doing the study itself tells you, 'hey, this method is likely to pick up changes that are induced by oxidation.' Certainly it is a useful tool to give you samples for your qualification study.

Deamidation is the other classic chemical degradant for proteins. Here we typically look at pH 9.0 in bicarbonate buffer. Sometimes you have to back off on the pH a little bit – typically over 24-hour time points. We like to follow this by both ion exchange as well as the isoquant kit. It is a commercially available kit. The reason we like this is that it measures essentially the isoaspartic acid content. It is not exclusive to deamidation. But if you do do a forced degradation study, one of the challenges you always have is, 'did it actually happen?' If your chromatogram does not change, the isoquant kit will in fact tell you whether or not deamidation occurred. With the oxidation you tend to have to run a peptide map or something like that.

Similar to the reverse-phase in oxidation, you basically are identifying the region of your chromatogram that will change. Once again, here is a 24-hour time course of a monoclonal antibody where the acidic species increases over time which ends up being singly deamidated product. Then you can see that over a further time course you can get even a multiply deamidated product. So once again, a relatively simple experiment takes a couple of days to execute and analyze and you know that this method is capable of being stability indicating.

The last example of a stress study that I want to go through is a **freeze/thaw**. This is not necessarily for your stability-indicating nature aspect, but it is really going to help you with sample handling. We like to do a variety of conditions and temperatures. Once again we track primarily by sizing as well as SDS-PAGE. Really what this helps you do: Number one, if you are going to freeze your bulk, it is nice to know this. But most importantly, this allows you to freeze your samples. Freezing samples can really help dampen the noise of your assay variability. Sample

handling ends up being very critical in early development. Being able to freeze samples gives you some flexibility in the inevitable assay downtime. And finally, it also allows you to have meaningful retains. Early on your assays may not be fully developed. It is nice to have samples left over from your earlier stability studies when a new assay comes up. You can actually generate several years of data if you have freeze/thaw availability.

Be careful with what conditions you look at. At –80 to 5°, which is the typical bulk freeze/thaw, we saw a very marginal increase in aggregate levels. Where we actually saw the biggest example was actually going from –80 to –20°. This is the formulation that did not have a cryoprotectant and had salt. You have essentially a PBS-type solution. And what is happening here is you detect it by the sodium chloride that causes freeze/thaw denaturation just above –20°. So as you go through that cycling period you can actually get a fair amount of aggregation....This tells you [that] if you have a frozen – 20° product, don't ship it on dry ice because it will go through that transition over and over.

Some other stress studies that I will mention here but won't go into detail on: **pH and heat**. These are both classical experiments, critical to doing formulation development. Actually, we find a little bit less utility in them when it comes to generating stability-indicating data because so many pathways happen with exposure to these conditions. Nevertheless they are still useful.

Light we typically don't perform in early clinical development for a variety of reasons. Mainly, the package protects the primary configuration from light generally.

Some **conclusions** from the stress studies: Hopefully, you have seen that certainly you verify that the assay detects the typical degradations. It will demonstrate your selectivity, which can certainly be used in qualification. One thing that is very important here is it will show the propensity of the molecule to degrade. Once again, going back to the risk-based assessment, [by looking at] more products, especially of the same class, you will be able to take your stress stability data and apply it into your database and see where on the pathway your product is likely to fall out of stability. And finally it helps you establish appropriate handling conditions.

CROMWELL: Stress studies are mentioned in several **ICH guidances** as I think has been highlighted throughout this meeting so far.

The goal of stress studies is to generate degraded product. The use of that is really to look at your analytical methods – to make sure you can pick up potential degradations that you are seeing, to figure out what degradation products are created so that if you have an extreme exposure in pH, what does that do to your protein? And finally, to determine whether those specific degradation products are process-related impurities or are they simply variants. To do this, we look at extremes of pH, oxidation, thermal stress and intense light, among others. So it is not limited to these that I have pointed out so far.

I am just going to walk through the guidances and highlight where it talks about stress studies....If you go to **Q1A** revision 2 section 2.1.2, it talks about stress testing. If you read into it, it says you should do it – include the effects of temperatures, humidity, which would be important for lyophilized products, look at a wide range of pH and do photostability testing.

Q2B: Validation of your analytical methods. Have stress samples for specificity purposes: light; heat; humidity; acid/base hydrolysis and oxidation.

Q1B is probably, in my opinion, one of the most detailed guidances that there is because it actually tells you what you need to do, which is refreshing. And on this, it takes you through a very clear decision tree. You start. You expose your product to light. Do you see a change? If you do see a change, is it acceptable? No? You put your product into its immediate packaging, which would be the vial with the label. Is it still seeing an acceptable change? No? Put it in the box. Do you still see a change? If you do see a change that is not acceptable, go back and start all over again with a redesign. Either you change your formulation [or] you change your packaging, so that you can eventually get over here to the 'end the test' because you do have an acceptable change.

Q5C, which is the biotech products guidance, also mentions stress conditions. Here it really goes into accidental exposures to conditions other than those proposed, such as those during shipping.

To highlight what is really covered in these, I think you have to look at the **selection of stress conditions**. It tells you that you need exposures to different pHs, you need light exposure, you need thermal exposure, you need oxidation. And what I would propose is that you not only do those, but you make them somewhat relevant to your drug substance production or what your drug substance will likely see.

The example here, with the **pH** extremes, is you can expose it to low pH, go down to pH 2, you can go up to pH 12, but you are probably going to see some very different things happening at those pHs, such as protein unfolding, that you would never, ever see during your manufacturing. In most processes, and here I will limit it to monoclonal antibodies, you are not going to go that far down where you would see protein unfolding. If you use the extremes for your manufacturing process, then you have somewhat of a basis to say, 'OK, these are real degradation products that I am likely to see if things go awry.'

Thermal exposure: You can get some information for that from your pharmaceutical development. For some proteins, going up to 50° is perfectly acceptable and you can do thermal exposure up there to get information. For a lot of proteins you will start unfolding and so 50° is not acceptable and you have to back off and find out a suitable temperature to do the stress condition that is not really considered accelerated stability. So you are walking somewhat of a fine line....Thermal exposure can be lack of heat as well. What if you freeze your sample? Do you see issues on freeze/thaw stability where you are generating aggregates [or] something like that?

Oxidation: You are probably going to need to experimentally determine what conditions it takes to oxidize your protein. We have had one case at Genentech where we have had a protein that we cannot oxidize to save our lives. We expose it to extreme levels of hydrogen peroxide, we see absolutely no changes on the peptide map. So the question actually became, well what do we do to the protein to make it oxidize? Do we keep throwing in more peroxide? Do we unfold the protein, make it oxidize so that we can show that our peptide map would pick it up if it were completely unfolded? I think, again, you could do that, but is it really relevant? Unfolded protein is not going to happen. If you do see unfolded protein, you are going to have other changes that you are picking up.

Since you can't go back to the point at which the material was being made for the clinic, he said, "what ends up happening is you have a lot of internal fights" about what to do with this information.

➢ CDER Monoclonal Antibodies Division biologist Joseph Kutza echoed the perspective expressed by his FDA colleagues during the CMC strategy forum that the problems created by the new assay information are more than offset by its value.

"Long term, my experience has been that the new assays, although they might cause some problem at first, end up giving you really good information," Kutza stated. "So I wouldn't hold back in developing new assays just to stay with the status quo."

Improving from a 10X to a 100X microscope will allow you to see more things, Ritter added. However, if these are "characteristics of the product that you expect, then there should be nothing wrong with that."

On the other hand, she recognized the challenge in trying to "dissect apart whether or not what you are seeing was actually present at the time it was in the clinic." She noted having seen study designs "where folks have to go back and try to determine what was the state of the product then." For "product-related impurities, the degradants that have occurred, there are studies you can do to force the stuff and prove that what you are seeing occurred over time as a degradation pathway as opposed to something that came through the process."

While challenges can present themselves, Ritter concurred with the FDA view that "it has never been a bad thing to get a more sensitive, more specific assay and then go back and sort of justify what you had back then." In this context, the retained samples become key. "With the retains you can have a case to build that it was or wasn't degraded, but without the retains it is all hypothetical," she said.

Assays Can Be Added And Subtracted

Among the concerns that received significant attention at the CMC forum was deciding how much of the assay work during development needed to be built into the early-stage or ongoing stability program and specifications and, in turn, how to choose which new assay tools to incorporate.

➢ Specifically, the participants were asked when selecting stability-indicating methods for early-stage stability studies if they "include all assays that show degradation during stress studies, even if they show no change at the intended storage condition," and if there are criteria for either removing or adding these assays later in development.

Shire Pharmaceuticals (formerly Transkaryotic Therapies) Pharmaceutical & Analytical Development Senior Director Zahra Shahrokh commented that "since the purpose of early studies is really getting information, I think it is safer to err on the side of collecting more data early on and then deciding which ones you want to drop later rather than not collecting because

you don't see a change in real time. And then again, you don't have a lot of real time data at that point."

Recognizing that new peaks will emerge as assays are improved, Shahrokh echoed other forum participants in stressing that "the key is having retained samples from tox and clinical material" to indicate if these are new species. She added that, in her experience, it is a frequent occurrence "that your assay could change and you could start seeing new things come out."

Genentech QC Clinical Development Director Wassim Nashabeh agreed that "bringing in new technology that shows us more information" should not be a big concern, "because especially throughout clinical development, we are at the stage where we are gathering information."

The purpose of stability testing at this stage, he pointed out, "is not to show no change, it is to manage the change in correspondence with the clinical program. So if we are two years into the program and now we have new tools that show us that we have something that we didn't see before, that is fine. The material has been at the clinics being tested. We have the data that correlates there. So I think we should encourage looking for tools that are sensitive [enough] to tell us what is going on with the product and shouldn't be too concerned about how will we deal with that information."

- CDER Division of Therapeutic Proteins Biochemistry Lab Chief Emily Shacter commented that "any assay that shows change in your product over time should be part of your stability protocol," adding that "there wouldn't really be any reason for leaving it out. Part of the point is to learn to see how that changed product might be impacting your clinical efficacy or safety, and not to test it would not be acceptable actually I think from the agency perspective."

FDA, Shacter affirmed, "certainly would not resist having any new clinical assay added. And I suppose there could also be circumstances where if you saw a certain change that was not in the tox lot, if you had and you tested your retain materials, there could be situations where there would be a requirement to…go back and do some tox studies to make sure there is nothing untoward in there that would pose a real safety risk."

Commenting on the issue of adding new methodologies to the stability program over time, MedImmune Analytical Biochemistry Senior Director Mark Schenerman agreed with Shacter and Nashabeh that "it is always beneficial to be looking at new technologies, especially technologies that are information rich that could give you more data about your molecule than the legacy methods."

However, he cautioned about the need to also be "responsible about understanding what the new technology data is saying." So that is why you have to collect data in parallel in stability studies and understand what the methods are telling you." He added that "characterization is an important part of that. Characterizing your molecule and your degradants early can give you a lot of confidence that the new technology is really something meaningful."

- In a dialogue later during the pre-marketing forum session on changing methods, Amgen Corporate Quality Compliance Associate Director Heather Simmerman injected a cautionary note, recommending that firms carefully evaluate the implications when choices are made on new approaches.

"I think we can over-engineer the methods and I think we have to ask again what is really needed to control the product," she said. "Just because you have a new assay or you have an assay that can resolve things" isn't definitive. "It really may be additional cost without an added value in terms of the impact to the product safety or efficacy."

In this regard, Simmerman suggested "a couple of things to keep in mind" when making the choices: A more precise assay will affect not just the specification range but also what is deemed significant in the data. Also with a tighter precision, "your system suitability ought to change in concert with that" which has a cost component. "You may simply be rejecting more assays without really positively impacting the quality of the product that is out there, and that is really not to anyone's benefit either."

A further issue is interpreting the meaning of the findings. More impurities may be seen, Simmerman said, but "are they relevant to the safety or the efficacy of the product? If not, then these improved methods certainly can be used for investigation purposes. And if you subsequently find out they are relevant, you can always add them. That would be your justification. But I think just because you can come up with something or a new technology, it doesn't mean that you should rush to implement it from a specification or a control standpoint."

- In summarizing the early stage discussions, Ritter supported the general point that firms should not be "afraid to collect the data during development so that you can provide for yourself and your reviewer and adequate amount of information on which to make sound decisions" about which assays add value and should "go forward and which assays should not."

If you have multiple assays that test the same degradation pathway and they don't provide any other orthogonal information on a different parameter, "then go with the one that is the most sensitive, because it is the one that is going to be the best for the stability indication," Ritter commented.

Lilly's Lewis also concurred with the viewpoint that characterization assays and assessment of "what your molecule might be doing under different situations" is important to carry through and "collect a body of knowledge about your product. And then if you have that information and you share that with the agency, you have a better opportunity to reduce what you might consider to be redundant assays and keep your sensitive assays in place."

Lewis stressed the value of carrying the characterization assays "all the way through up to licensure and even beyond in some cases with regard to comparability. Even if you don't have them under the stability protocol, make sure that those assays are available" later to evaluate changes to the process or container/closure.

Role of Potency Assays Debated

The role of potency assays in the stability evaluations was a particular focus of attention at the forum.

During the late clinical phase discussions, Shacter commented that "whether or not your cell-based potency assay is stability indicating or not, it should be a part of your stability protocol. There isn't really any case where we would not want to see the potency assay done at every time point." The potency assays "are not always stability indicating, but they are one of the bottom-line parameters of a product. We need to know how anything that you see might or might not be impacting potency."

In support of Shacter's point, Genentech's Nashabeh noted that depending on the development or clinical stage, "you might not have enough confidence to determine whether something affects potency or not because you are evolving in your selection of your potency tests, and I think you need to monitor anything that changes until you gather that body of data prior to going to commercial."

- Addressing the incorporation of stress studies, Nashabeh clarified that Genentech does not view oxidation as "one of the standard tests that is on a stability protocol or release, but it can be added if the product has a certain susceptibility to oxidation based on the data we gather."

Diosynth Biotechnology Customer Project Management Director Siddharth Advant affirmed that stress stability studies "are very important" in giving information about the molecule. He queried the audience whether they repeated the stress studies as multiple lots are made through clinical development – for example, on changing the scale.

"Ultimately you want to go after real-time stability data," Advant said. "But if you start looking at some of these profiles from stress stability studies, especially temperature profiles, and if you see that multiple lots made at different times show you the same rates of degradation for example, that could give you maybe some level of confidence that at a year or year and a half, you probably are going to see the same degradation profile in your real time."

Nashabeh concurred that "especially when you go through process changes this may be one of your comparability parameters that you look at because you are not going to wait a few years to get real-time data." These stress studies, he pointed out, "are not perfect. There are a lot of unknowns with them, and the rates may not be as predictable as what you would get in the real-time studies. But it is one tool that you use as long as you use it in context."

- Ritter added to Nashabeh's point by noting that the value of the forced degradation "goes side by side in a comparability protocol….If these things are truly comparable they should degrade with the same kinetics and make the same degradants under the same physical or chemical stresses."

She has seen "a lot of examples where people utilize that and then gather that data at the end of their product development lifecycle and present it as a nice package to say 'not only does this demonstrate comparability from change to change to change, but over time it also supports the stability-indicating methods. It supports the

degradation pathways that we have mapped out. It supports the key degradants of the product.' And it makes a really nice story when you have got it at the end."

Division of Therapeutic Proteins Deputy Director Barry Cherney pointed out that the concept of using stress studies in assessing comparability is incorporated into the ICH biotech comparability document Q5E.

- A concern was brought up at this point by Amgen's Simmerman about "the use of the concept of comparability or comparability protocols for the early development work." The concept, she said, is really intended to apply to the product in its more mature, commercial form.

During the development cycle, Simmerman stressed, "we intend to make changes and frankly we don't want things to look the same. We want to be making them better." In turn, having protocols with prospective acceptance criteria, she cautioned, is not "entirely appropriate for the early development phases." While firms want to do comparisons and determine differences at this stage, "we need to expect the differences and in fact look for those improvements…. But the use of comparability, comparability protocols, the notion of acceptance criteria that in fact they are the same, I don't think that is really what we are intending here at the development stages."

Lilly Principle Regulatory Scientist John Dougherty agreed that a comparability protocol is primarily a post-approval regulatory instrument, suggesting that terminology such as "bridging studies" is more applicable in development.

Simmerman reiterated her concern that "the clarification of the terminology and the expectations is very important, because otherwise it is looking to me like we are setting a new standard and expectation in clinical development, and I don't think it quite works….The terminology for clinical development does need to be differentiated from the post-approval stage."

She noted that Amgen has "made some pretty dramatic changes in clinical development cell lines, cell sources, all these kinds of things, and they look very different from a profile perspective. They look very different from a stability perspective as you expect. Yes there was a bridging study, yes there were comparisons made, but it wasn't framed as a comparability. The verbiage I think that was used was, 'this was equivalent and suitable to go forward with.'"

Office of New Drug Chemistry Team Leader Stephen Moore noted FDA's use of the term "linked" in relationship with toxicology or early clinical trial material.

On the other hand, fellow CDER official Cherney pointed out that Q5E, while recognizing the concerns Simmerman raised, "does talk about comparability during development." He reassured participants that the agency understands the nuances of the comparability concept and its different application in development and after pivotal trials.

Determining Safety/Efficacy Impact Is The Holy Grail

Genentech QC official Ruzica Djerki reframed the discussion of early-stage concerns into two components: how stability is monitored, and then how the information should be interpreted.

The monitoring, she said, is relatively straight-forward. "Basically we all do the maximum we can," with the main reliance on concurrent testing. On the other hand, "the real challenge is…when we see the changes, how and what are the tolerances there – what are the acceptances for changes that we see at that early stage.

The struggle, Djerki said, is around "what we are sending to our clinic. Should we send the worst materials so that we have more space there? How we can be sure that all our assays are covering all these things, that we are on the safe side and that certain things are tolerable – that is much bigger question for me." It is not so much "how we do the monitoring, but when we see the changes in monitoring, how we are addressing those."

- The quandaries encountered when attempting to correlate stability and other CMC development information to clinical safety and efficacy have been receiving significant attention at FDA/industry conferences over the past few years ("The Gold Sheet" February 2004), and the issues were wrestled with again at the stability strategy forum and the follow-up workshop.

Reviewing the forum discussions on the issue of clinical relevance at the workshop, Ritter commented that "trying to link the product attributes, especially a stability attribute, to whether it is going to remain safe and efficacious in the clinic…is of course the Holy Grail of development."

Stability Study Design Considerations For Late-Phase Development Programs

The following overview of late-phase stability study design considerations was presented by CDER Division of Monoclonal Antibodies Biologist Michelle Frazier-Jessen at the recent CMC Forum held at NIH. Along with general requirements, she discussed reduced stability study designs, accelerated and stress studies, and bridging studies for changes in manufacturing, formulation, container closures, and analytical assays.

I don't know how many of you have looked at the draft guidance for ICH Q8 – it is on pharmaceutical development. I like the document. I think that it has a lot interesting points in it that I look at when I am thinking about a product as it is going through development stages. Really, the aim of any pharmaceutical development program is to deliver a good quality product and to design a manufacturing process to deliver that product in a reproducible manner. Any information that you gain along the way really helps to support what you know about the product and to be able to control manufacturing process and to set specifications.

The way that I am thinking about this is more along the lines of late-phase development. I am thinking late Phase 2, pre-Phase 3 to the submission of the BLA. I am just going to give you a brief outline. I am going to talk about: • general requirements • some interpretations of what is meant by representative • reduced stability study designs • accelerated and stressed stability studies, and • bridging studies.

When I think about the late phase program, I am really thinking at this point [about] a product. There should be enough information that has been generated in the early studies to demonstrate how the product is going to degrade, what pathways are going to be utilized. That should be fairly well understood at this point in time. I mean, certainly you can never know everything, but you should have a good idea of how your product is going to degrade under its proposed storage conditions and packaging, etc.

You should have a set of analytical assays that provide a stability-indicating profile. I feel that potency needs to be part of that profile. A potency assay may not be the most sensitive stability-indicating assay that you have in your little package, but it should be able to be stability-indicating. If it is not I think that is sort of a problem, in my mind.

Additionally, you should be well into your stability program that you are going to use to support the licensure of your product. That should be ongoing. You also might be making some changes to your product to prepare it for licensure that might alter how you do your stability program. You might be scaling up. You might be making some formulation changes or container closure changes, things like that. But you should be collecting data on all of these things at this point in time for stability.

General Requirements

Looking at the general stability requirements [in ICH Q1A and Q5C] for drug substances and drug product, you obviously need real time and you need to have some accelerated. That would help support excursions that you might have for shipping or storage or any kind of things like that. And you need at least three batches. They can be full scale or pilot scale. This is what the guidance says.

Pilot scale needs to be representative of the manufacturing process. If for some reason, and it says this in the guidance, you were to use pilot scale, you could use pilot scale for a license application with a commitment to place your first three full-scale batches into long-term stability program after approval. I don't know that that is necessarily the practice…for the obvious reason of the risk that entails. I have not seen it but others might have.

You need to have defined the **storage conditions** that you are operating under: The relative thermal stability of your product. You should be evaluating the sensitivity to moisture, especially if it is a lyophilized product. This should be done in a representative container closure system.

I think you need to do some work with your **intermediates** as far as stability goes. You need to identify them. You need to generate some in-house data on them and the process limits that you are going to operate around. Although the guidances say…it might be possible with justification to use representative scale or pilot scale, you really should be establishing any kind of stability specifications for intermediates at the validated manufacturing scale process for obvious reasons. This needs to be real-time, real-condition data and will likely require more than one batch.

What is Representative?

I think we should get into the term 'representative.' I think…there are some vagaries here [on] how 'representative' is interpreted: I think you can talk about small scale and you can talk about pilot scale.

For small scale: If you are using small scale for drug substance stability, if you look at Q5C, you can use reduced size containers. That might be acceptable for storage. But the drug substance that you use should be stored in containers and the containers need to properly represent the actual holding containers used during manufacture. So you need to justify that those containers are properly representative.

With regard to pilot scale, as we had mentioned earlier you can use a pilot scale batch with proper justification. Obviously it needs to be fully representative of, and simulating that, to be applied to the full production scale batch. It does not really define a number or anything like that. It is interesting because the EMEA has further defined this and said that it needs to be at least 10% of the production scale batch. But it is not written that way in Q1, at least not to my knowledge, though maybe it has changed now – I don't know for sure. The additional thing with pilot scale that people need to remember is that it needs to be operating under current GMPs. I think people sometimes think 'pilot scale' and they forget it still needs to be under current GMPs...

Reduced Stability Study Designs

With regard to reduced stability study designs: Certainly at this point in time, while you are doing your development, you might be considering different dosage forms or dosage sizes and things like that. It is possible to apply some reduced study designs. It is usually for drug products, but that might not be acceptable if you have a very complex drug delivery system, so that might not be the appropriate thing to do. As I said, it is going to require justification. You are going to have to have supporting data. The supporting data are going to have to show that you don't have too much variability and the product is pretty stable.

Two common methods that are used are **bracketing and matrixing**. Bracketing is really a design of a schedule such that you only test extremes of certain design factors. This is pretty common… if you have different strengths. I will show you an example in a minute. It is generally not applicable for drug substances. It is mainly used for drug products.

[This] is taken from Q1D. You have 50 mg, 75 mg and a 100 mg. There are three different batches and you are just testing the extremes – the 15 ml size and the 500 ml size. The important thing to remember about this is that if you should decide to drop out one of those extremes, for example, like the container size or one of the dosages – you are not going to go through licensure with that, you decide not to carry it on – if you are doing your bracketing study and you have set that up you are still going to have to carry that bracketing study through the licensure as a post-marketing commitment so that we have that data, because that is how the study was initially set up. I think that it is a good idea that if these are the kinds of things you are going to be doing and you are going to be submitting to the agency, that when you have a pre-BLA meeting you bring these things up and make sure that what you are doing is going to be okay with the review group.

Matrixing is a little bit different. You are looking at a selected subset of the total number of possible samples. It is okay if your supporting data indicate that you don't have a lot of variability in your product. The more variability you have, the more justification you are going to need to be able to do a design like this. And it is really not very useful for drug substances.

This is an example also taken from Q1D. You can see this is a one-half reduction, so you are only testing about half of the time points. It is a matrix. You are doing a little bit here and a little bit there. That might be applicable. And once again, if you are going to do this, you need during your pre-BLA meeting – which I encourage everyone to have – you should be discussing these things and seeing if they are going to be acceptable.

Accelerated and Stress Stability Studies

I saw a lot of questions before [at the CMC strategy forum] that were put out regarding the accelerated and stress stability studies and what they are and when do they need to be done or why should they be done. So I thought it would be useful to talk about this a little bit more.

Accelerated testing: It is really studies that are designed to increase the rate of chemical degradation by using exaggerated storage conditions. Stress testing is a little bit more than that. You are really trying to elucidate the intrinsic ability of the product to degrade and it is really carried out under much more severe conditions than you would use for accelerated testing. With accelerated testing, obviously I think most people think of accelerated temperatures and things like that. Stress testing can run the gamut of agitation, pH extremes, oxidation, things like that. And for drug product, if you are going to do stress testing, another one that is very common is to look at photostability.

For stressing your samples: This is really quite useful to look at different degradation pathways, and it can help you to validate different methods. I guess probably an example that I think of most commonly is…if you want to change an analytical assay out, how do you know that that assay is going to be an improvement over the initial one? One way to go about looking at that is doing some stress stability studies and determining if your new method is at least as sensitive or maybe more sensitive to picking up these stress conditions, the degradation of the product.

According to Q1A, it may not be necessary to do some of these. I think what it really depends upon is how well you understand your product and how it degrades and the justification of what you need to do based upon that knowledge. So the more knowledge you have about how your product degrades – which I would think at this point in time of development, you would know that – then that will define the kinds of stress studies that you would do.

Bridging Studies

Then bridging. Oftentimes at this point in time a lot of sponsors are making a lot of changes to be able to bring their product to licensure. They can be manufacturing changes to scale up, to remove impurities, things like that, formulation changes, changes in the container closure system or changes as I just mentioned in analytical method or methods.

Why would you make **changes in manufacturing**? Well, you might be improving your manufacturing process. You might be increasing the scale, which is extremely common right before Phase 3, and you want to improve your product stability because you have gained more experience with the product and you understand how to keep it more stable. Or you might have to comply with changes in regulatory requirements.

For changes in manufacturing, certain changes, even though they might seem slight, have the potential to really alter the stability of a product. Certainly any change that can alter your protein structure or purity and impurity profiles really needs to be evaluated – not just on real-time, but under accelerated and stress conditions. And not only just your drug substance and your drug product but also your intermediates. I think people often forget that and sometimes that is where you have problems.

Stability studies that are undertaken can also be really helpful if you have a subtle difference and it just may not be detectible by the characterization studies that you have. So certainly, long-term stability studies and even accelerated or stressed under these conditions, comparing the two different processes, you might pick something up. One of the common ones that comes to mind...[are] proteases and divalent ions. But certainly, even a slight alteration in a protease level that may not necessarily be picked up because of the sensitivity of your assay, or it is so small it just does not seem like much, might actually really cause a difference in the degradation profile or the pattern or just the amount of degradation that you get. You may not pick it up unless you have done the real-time studies and some of the accelerated studies. I have certainly seen that happen where we have had a change in manufacturing process and very subtle differences in the purity levels – and we are still monitoring this product, obviously, because it is fairly early on – but you do see a difference in the impurity patterns. So it does happen.

Formulation is another one. Usually I think, by this time, most people…start out with an initial formulation and during development, you might make some changes, certainly to improve the stability. But at later stages, it might also be because you might be wanting to use a new administration route or delivery for the product. I think everybody knows that not all formulations are created equal. Every product is different and what works for one product might not work for the other.

This is just a good example of what happens. This is accelerated stability of a monoclonal antibody. This is an AUC data and you can see that the buffers,...not all of them are able to keep the product stable. If you are going to be changing your buffers you need to not only evaluate them under the normal storage conditions but also under stress conditions because you might have to do more work with regard to that.

I can think of a recent example that I just had. I had a sponsor that has a monoclonal antibody under IND, and they decided to go after a new clinical indication and this clinical indication required a different administration route. That administration route was not compatible with the current formulation scheme. So they re-formulated their monoclonal antibody into a phosphate-buffered saline buffer and initiated preclinical studies to support this new route. Then because it was going to be at a couple of different clinics, they decided the best thing to do would be to store it at –20.

I don't know how many people know much about phosphate-buffered saline, or phosphate buffers and protein products, but PBS-based buffers tend to form protein adducts or you get pH gradients. So this product dramatically degraded upon storage. They were using it and not realizing it for their preclinical studies. It turned out I think that they were okay because the very high concentration that they had actually was fairly stable, but the other two concentrations degraded dramatically, and they did not run their stability studies until after they had finished their pre-clinical studies. This could have been a really big ouch. I think it was a little bit of an ouch, but certainly was very painful.

The other big thing I think that a lot of sponsors do at this point in time is changing their **container closure system**. Maybe they have been in a vial and they have decided they want to go to a pre-filled syringe. I think that that is becoming much more common, that choice or that pathway. Certainly a new administration route might be a reason to do that. Certainly if you are going to something like a pre-filled syringe format, that is much easier for the physician and the patient to use so that would be of benefit. You might also have to change your container closure system because what you were using before is not available or you have decided you want to make a couple different versions – not just a 100 ml vial but maybe a 150 ml vial as well.

These are the kinds of studies that tend to lend themselves well to the bracketing studies. So it is important to remember with container closure systems that they are suitable for the product that you are using, and they are suitable for the intended storage and transportation, and that you have evaluated the potential interaction between all the contact areas. Extractables and leachates are things that you need to be looking at with these kinds of things, especially for the pre-filled syringes.

Q1A says you need to have stability data in the to-be-marketed format or something that looks, that is representative – once again the 'representative' word. I think that for some of these novel or newer container closure systems that we don't have as much experience with, to me that means that it needs to be in that container closure system. So it would be hard to go from a vial and say that that was an acceptable bridging. You really need to have some data in your container closure system. And you should probably have some clinical experience with that as well.

One of the big things that I certainly see happening a lot is the change in an **analytical assay**. You are replacing an existing method with hopefully an improved method. That would be the reason. This is extremely common for monoclonal antibodies with regards to the potency assay at this stage. A lot of times what sponsors come in with in the initial IND is...just a binding assay for potency. While that is allowed, sponsors are reminded that that might not be acceptable for licensure. Usually during the course of development, at this point in time, a new potency assay that is representative of the biological mechanism of action hopefully is coming on board. And that assay will need to be transitioned into the stability study, and it should be stability-indicating to some degree.

Certainly this is not all of the data that might be required to support, but just an example of what you might need to do with regard to an analytical assay that you are changing out, for example for potency. You want to look at not just the normal conditions, but you want to look at some stressed and accelerated data and compare the assays and see which assay really performs better, is more sensitive to the different degradation pathways as you understand them and is not as variable. And you want to compare this data in the context of your other known stability-indicating assays. So there should be some kind of a correlation to get an idea of what is going on. That would be a good reason to change out to a new assay.

The big issue really is **developing a quality product** and really understanding your manufacturing process and how your product degrades. And understanding really the big picture of 'this is the indication, this is how the drug is going to be delivered, stored, transported, etc.' All those things taken into consideration and all the data that you have generated will help support your application and help you to derive your specifications and shelf-life and things like that. Any changes that you encounter, or any data that you encounter, even negative data, are extremely useful from a regulatory perspective because it allows us to get an insight into the kind of knowledge you have on your product and make good decisions with regards to these important things like specifications and shelf-life.

Reflecting the discussions, she pointed out the basic dilemma that arises in this pursuit: "The most empirical way to figure out if there is a link is to take distressed material into the clinic. But that isn't necessarily a particularly sound strategy for a variety of reasons, one of which is that you don't want to do anything to jeopardize your clinical trials – not just the trial data, but in fact the patients themselves."

Ritter noted that participants at the forum did not indicate that there was an effort to date to use aged/degraded material in a systematic way in the clinic, although retrospective monitoring of the age of the samples and the product profile when used was an approach that was being deployed to help assess clinical correlations.

➢ Lilly's Lewis sees a "Catch-22" in FDA's emphasis on knowing the impurity profiles of the clinical trial materials, with the idea that data may then be available to potentially support higher levels.

"The problem is that it is a slippery slope, at least in my mind, in terms of …putting degraded material into the clinic, or even holding back some aged material and putting that into the clinic. How far do you go?" In turn, "especially early phase, there is not enough power in those studies to be able to discern whether there is any differences between a certain age material and fresh material. That is a big issue."

Lewis pointed to the idea he hears expressed that "if there is not adverse events that we find with aged material, then we are okay." However, he sees the need for "more discussion on that topic," suggesting that "it is somewhat dangerous to design these, and especially in an early phase, even if your toxicology data may support higher levels."

Ritter concurred. "One of the burdens that we bear in biotech," she said, "is not just that the loss of our active could cause a loss of efficacy in the product, but clearly the fact that our degradants are themselves potentially harmful in terms of immunogenicity. And that is not something which can easily be discerned in a short-term study or a small sample size."

Ritter pointed out that this concern is another reason for continuing with the full regimen of assays during development. Doing so allows for retrospective analysis of the batches "and actually matching up in some sort of a matrix – what was the age of the material, what was the actual state of the material at the times it was used in the clinic, and correlating that with clinical use – not necessarily in terms of being able to look at long-term safety problems, but at least in the short term being able to say that these are characteristics of the material that have been subjected to clinical trials."

The problem, she noted, is that such an approach requires "a cross-disciplined discussion, and sometimes it is very difficult to get the clinical people and the CMC people together with the right data sets to be able to overlay them and make some sense of it. But that is really what it takes to be able to match this stuff up."

EU Clinical Supply Expiry Mandate Raises Concerns

Another key focus of attention during the early-stage discussion was on the issues involved in stability determinations and dating for clinical supplies, given limited experience with the product, manufacturing and methodologies.

In her summary of the forum discussions on the implications of the limited manufacturing base and lack of knowledge of batch-to-batch variability, Ritter noted general agreement that "you always build on your experience gained from development lots." She advised firms to "try to draw on that stuff. I know that it is very difficult sometimes when a project is moved from hand to hand to hand, but that information is a very important part of the knowledge of the product, even very early on."

Noting that "we don't have a lot published right now in the biotech community about a wide variety of molecules," Ritter again stressed the importance of sequestering materials from those early development activities and "using that to build on your knowledge and be able to start to infer some of the known degradation pathways for molecular types."

In his presentation on early-stage stability assessments, Human Genome Sciences' Spitznagel also noted that "the early development studies are going to provide the most data...and you certainly want to use it." In turn, "the better you set [the studies] up, the more applicable the data is going to be." He emphasized the importance of stress studies, in particular, as "probably the most critical aspect of a good development program" (*see box, pp. 4-6*).

In general, Spitznagel concluded, in setting clinical expiry or retest dates, the basic point is that "no single approach will work in all cases." The key, he asserted, is knowing the propensity of the product to degrade, and

then "always monitoring your clinical lot concurrently." Extrapolation is "something you certainly should do. It can certainly be comforting and can actually provide some very good justification for setting an expiry period. But you have got to be real careful with the actual predictive numerical value that comes out of there. It is often not going to give you the actual expiry period."

- The issues around setting stability parameters for clinical trial supplies are receiving heightened attention in the wake of the new EU clinical trial directive, which calls for all clinical supplies to be labeled with an expiration date rather than the practice in the U.S. of retest dates. Participants at the CMC forum noted that the requirement is forcing firms to do extrapolations in spite of their limitations.

Ritter commented that the EU requirement is "logistically very difficult. Apart from the extrapolation or how you get to that, the logistics of physically labeling the stuff has been problematic."

Firms with European trials, she said, are "deferring right now to ICH Q1E," but it is "very difficult" to apply the guidance to extrapolations for biotech products, particularly in terms of correlating the different assays involved. The decision for some firms, Ritter explained, is to say, "it is frozen, we are going to put a year on there. We will test it. If it fails, we will pull it out of the clinic."

CDER's Shacter commented that from FDA's "internal perspective, it is more important" to use real-time evaluations to "stay ahead of the curve…so you know what to expect with the stability of your product while you are in early clinical trials…because we understand that there isn't a good basis for having an expiration date per se."

CDER biotechnology office colleague Elana Gubina also expressed the need for caution. She noted having seen examples where "these extrapolations, especially based on a not-well-validated potency assay, can lead to very serious consequences." Extrapolation "may be a very useful tool, but it is a tool. Real-time stability data so far are much more valuable."

- Gubina joined Spitznagel in affirming the important role played by the early-stage stability studies.

Noting that Phase 2 trials are based on the dose chosen from Phase 1, Gubina pointed out that "if your product is not very stable during development, you really don't know what to base your dose on."

Spitznagel maintained that the expiry period is actually "one of the lower uses of a stability program in early development." Confirming that the product is within specs "is important from a safety perspective" and is "what you are going to file with the agency." However, he noted, "from a development perspective," the stability work "is giving you that lot-to-lot variability data. It is giving you data about your assays" and allowing "you to learn a fair amount about your product during the early clinical program. So even if it doesn't necessarily predict the expiry period beautifully, you are going to get a lot of scientific information out of it."

- One of the focal points at the late-stage clinical/application session of the CMC forum was the use of small-scale studies to support the stability of the drug substance.

Participants agreed that small scale studies are a common and acceptable practice if the small scale is sufficiently representative. "Representive," Genentech's Nashabeh commented in summarizing the discussions, "means that you take into account material of construction and other technical issues regarding headspace and so on, and that you will preferably have a justification in terms of how you have made that selection."

Participants in the small-scale dialogue pointed out that small scale often represents a worst case situation, compared to large scale where there is less manual manipulation.

Post-Approval Testing Regimen Changes Addressed

"Let's assume that you have gone through your clinical development profile. You have done all these studies. You have shown that some of these assays are not stability indicating. At what point post marketing," Advant queried, "can you actually take them off? Is there precedence from the industry that they have been able to take them off?" A related question, he said, is "can you take off some time points if you are doing, you know, zero, three, six, nine, twelve? Can you do minimal time points if you have shown that the product is stable?"

The debate over the criteria for adding, replacing or deleting assays in the stability program continued in the post-marketing session of the forum in response to a prepared question presented to the participants by Diosynth's Advant, who was serving as moderator.

Amgen's Hasselbacher On Developing A Comprehensive Stability Plan

At the CMC Strategy Forum on stability issues for biotech products held at NIH in late July, Amgen Head of Corporate Stability Carol Hasselbacher gave the following presentation on the "Development of a Comprehensive Stability Program for Product Commercialization."

I am going to continue with our talk on late-phase stability programs by starting with an example. Amgen is a company that has been fortunate enough to have a number of licensed products on the market, and I know that a lot of those of you that I have talked to this morning and others still have products at your companies that are solely in development. So what I was asked to talk about today was what it looks like from the other side. What does it look like after you are through with process and product commercialization? What have you learned? So the questions that I am going to try to ask are, "what could you do better?" and "what did we learn?"

One thing that I have learned is that there is a difference between a stability program and a stability strategy. The **stability program**: We have all been talking about the requirements….These are things that are the 'how.' This is how you do it. It includes tests, time points, lots on study, special studies, etc. And to succeed, you first of all need to go through the guidance….We also need to get feedback from the agency, and we are provided with lots of opportunities to do that during development. We also need to benchmark industry standards because there are lots of times when the requirements are evolving or expectations are evolving. And just by sitting at your own company and looking at how you have always done it, you are not going to be aware of these new standards.

But what I've learned is that we also need to include a **stability strategy**, which is the why you do what you do. And the question to ask here is, 'how does your stability program, how does the aggregate group of protocols that you have for your product, support your product commercialization goals?'

This is just a default product development timeline. I am sure a lot of you have variations of this at your company. The arrows are familiar developmental stages from early target through Phase 1, Phase 2, Phase 3, filing and launch. If you look up above where that large yellow arrow is, that is your stability program areas as I see it. You have got pre-clinical, Phase 1 and Phase 2. And then our focus is going to be on primary stability, the primary stability process, and getting ready for commercial. But these are all parts of the program.

If you look down below that group of arrows, you see a number of portals. For example, first-in-human portal, where you decide is the product ready to go into the clinic, up to end-of-Phase 2 portal – are we ready for our pivotal studies? Commit to launch – are we ready to launch? Most companies go through this sort of staged assessment, either formally or informally. But these have more to do with your stability strategy because at each point, you want to make sure that your stability programs are ready for the product development goals.

Here is an example, and this is something that we have talked about a little bit this morning with respect to early development. For a stability program, early development, your first and only GMP lot is put on stability. This meets all your requirements. That is fine, that is your program. But what about if you would be able to manufacture a couple of smaller lots and put them on stability. What does that require? That requires a lot more work up front, probably. It requires persuasion. It requires a good participation for the stability group within the project development plan and teamwork across functions. And it also requires for you probably to have some sort of documented rationale for what you hope to gain from this. But then you end up with clinical experience that is not limited to one data point. So this is sort of the back and forth that I am talking about between work up front and fast to the clinic, and where the stability groups need to position themselves as advocates in this process.

Of course stability strategies are phase-dependent. Each stage of development has its milestones and endpoints. But nobody knows when a product development program is kicked off where you are going to end up, what the product will look like. There are too many changes along the way. So what do you do about these changes? Stability has a big problem because our work depends on real time. We can't throw resources at a problem and have it go away. We need to sit tight and let the stability samples incubate for a certain number of months. So sometimes it is helpful to just think of things more generally in terms of design phase, which would be research and pre-clinical, development phase which would accompany the body of the development work, and confirmation phase with your primary end and conformance batches.

Have Commercialization in Mind

For those of you who haven't been to this point yet, I would say always start with commercialization in mind....Think first about the commercial phase and how to realize your development objectives. Work with the product teams also to determine timing and prioritization for each requirement. And you also need to do good risk assessment. You need to define what the risks are and communicate those risks to the project team when you can't meet your timelines or when there is work that you would advocate that can't get done.

So what are the **goals** for commercialization? I can think of three right off hand: ● one of them would be develop appropriate stability-indicating assays ● also hand in hand with this is develop a thorough product understanding, and as I have been talking about earlier ● understand and support the product deliverables.

First of all, whether you started out with a few assays and were adding assays as you went, or you started out with a whole bucket of assays and you are refining them, you want to be at a point in the late clinical phase where your assays are appropriate and comprehensive, and each test had better have a discreet purpose. You don't want a lot of redundant testing going on. You also need to document your rationale for this panel of tests. Once you are gone and there is no record of why this set of stability tests was chosen, it is not going to be very good for future product work so you need to document everything.

Hand in hand with this you need to: document your physical/chemical characterization data and summaries; define your degradation pathways;...also establish and document the stability profile. A word about documentation. It is really important. You really have to have a system at your company for archiving information, making it easily accessible to others. This is very hard to do in real time, but it is also much harder to do retrospectively as I am sure you all know. So it is something that just needs to get done.

For analytical considerations, the first goal in the design phase…early formulation development. It is often not the commercial development. Commercial formulation and stability studies are minimal at this point. They are used to support the tox lots and recommend an early formulation. The tests as was mentioned this morning are also quite general, probably not optimized. If you are working on monoclonals you may have platform strategies that you use at this point.

During development: This is where analytical formulation, process, characterization all go on in tandem. Your assays are being developed in parallel with understanding the process and the product. So at the end you have to have robust, appropriate assays, and the commercial process, formulation and presentation must be defined.

Confirmation phase: You want to have your assays validated, tailored for the product, and suitable for monitoring quality and consistency.

For characterization, this is going on at the same time. Design phase: very little happening in the PD and quality world. A lot going on in research. In the development phase, though, you begin to understand the product degradation pathways and obtain this understanding physical/chemical analysis, theoretical deamidation sites – look for that in the primary structure, stress studies. But along with this, your stability experience is what is going to provide your practical understanding of what really happens with the product. You don't really care about some stress situation that will never happen with your product except that you want to be able to make sure that it doesn't happen. Your real-time studies and your accelerated studies are going to be very important here. Confirmation phase: product fully characterized – this is not the time to do development.

Some **deliverables** are covered in the stability program plan; for instance, the ICH guidance, other regulatory requirements. We all know that we can go to the guidance and find out when to do photostability studies, for instance, and how to do them. But some deliverables are going to be linked to your specific product strategy and very dependent on your individual product plan. So you have to do teamwork to understand what those project goals are.

As an example, let us say you are going to do a global Phase 3 study. How do you support clinical excursions? You may have regions where you have never been before in the clinic and these regions may not have appropriate storage units. What do you do? How do you support that? Likewise with transport. What happens if you have a study that is ongoing in Brazil and the product is sitting in a warehouse for a day? You don't want to be constantly throwing away your clinical product, so you have to be able to develop more robust systems along the way if you are going global.

How to meet regional expectations? I'll talk about this a little bit later. This is something that has been a big issue for a number of companies recently.

General Strategies and Complications

So some **general strategies**: Develop adequate clinical experience – goes without saying. Try to use a set of batches in clinical development with a range of attributes. Also reserve samples – we have heard that already this morning and that is an excellent idea. Demonstrate that the product can be held frozen and then freeze samples and store them because the assays will not be the same later in development.

Define your commercial manufacturing process and plan presentation as early as possible. This is not always easy because marketing is always waiting to see what people want and often you will find changes towards the end of development. But try to be an advocate for getting that information early. Design space…is something that is very important during development and the stability function can advocate for this. The design space idea is intended for process development. However, I think it does have application to stability.

Quality by design…involves starting out with the idea that you are going to build quality in from the beginning. You document what is good enough, what is acceptable, and then you develop your product according to those deliverables. This concept may be applicable to stability simply because the more you know about your product, the more flexibility you are going to have at commercialization.

Product understanding and assay development also go hand in hand, as I have said. The goal is to gain an understanding of the product so that chosen tests monitor the relevant stability parameters. That is a big issue, probably one of the biggest ones. By the time you are at commercialization, you want to be sure that your tests are robust and that they are complete and that they monitor everything that should be monitored for your product.

So all of these ideas – if you can be an advocate for your group and help make sure that this development happens at an early time, you will end up with a better assay profile, better ways to do out-of-trend analysis, how to resolve nonconformances, and you will be able to bracket and matrix more easily.

Of course there are **complications**….Those of us who are in a business do understand that there is always change going on with development timelines, priorities, resourcing issues, groups being moved from one project to another. The idea of speed to the clinic, while it is essential, means often that there is less up front characterization and less method development up front. There are always changes in formulation, presentation, manufacturing process. And often you have a small number of bulk clinical batches so you don't have a very good understanding of the variability of your process early on.

The results of this are that your product knowledge at commercialization may be impacted and you may lack needed product understanding and expertise limits. Your options will be limited and your opportunities as well at commercialization. This is just a word to those of you who haven't been there – keep looking forward. The reason is remediation and catch-up after licensure is both risky and very expensive.

So what do you do when your project team says, 'sorry but we have to go into the clinic in two months?' You start with an understanding of what you want to have at filing. Always keep that in your mind. You actively participate on the product and project teams. You drive for needed development data and tests and escalate issues.

However, you need to use resources wisely. You can't gain credibility by asking for studies that are unnecessary or irrelevant. So you really need to understand what the issues are. Often, at least in larger companies, you will find that there are other groups that are doing comparable work, so you can actually find information that you may not know existed.

Think proactively. It is easy for those of us managing stability programs to put samples aside and use them for a sample library for later on because we know that for comparability we are going to need them.

Global Considerations

I am going to talk a little bit about global considerations now. ICH harmonization really provides a lot of value for those of us who are trying to understand what countries require. One of the things that it is hard for people to understand is that good stability programs may not ensure good global strategy for stability. The

reasons for this – I think there are two reasons. One is that while the small molecules guidance is fairly prescriptive, the large molecule guidance is open to interpretation, as it has to be, and some countries interpret things differently. Also, there are regional or customary expectations and traditions that predated the adoption of the guidance or are otherwise still very ingrained in a culture in another country.

When you develop a program, you need to accommodate regional requirements to allow for business model changes, because you don't always know where you're going to be marketing your product or where you are going to want to apply for licensure later. For stability, the timeline and cost issues for doing additional studies is untenable most of the time.

A couple of things to add to this: The FDA makes it quite easy through a lot of possibilities for interaction throughout development for companies to come to the agency and say, 'you know, we are making these changes.' Or, 'this is our strategy, is this acceptable?' Other countries don't do this. I am thinking particularly of **Japan** where between the time that you file your application for a clinical trial and the time that you file your application for licensure, there is no contact with the agency. So all of that bulk of years of development work is invisible to the regulators, and they only see it when you file your license application. So your question has to be, 'were these changes acceptable? What changes can we make?'

There is another issue and that is the annual report structure that the US has is very helpful. And I can think of this in contrast to the European Union which doesn't have such a system. Where, for instance, for a process change maybe you would need six months data in Europe for the bulk, you might get away with one month data for the US, because the US knows they are going to be getting updates periodically. So there are good reasons why there are different requirements.

So for Japan, how do you succeed? First, you need to accept that regulatory requirements are different…. It is hard for some companies to believe that there is really a difference and so a lot of it is just getting over that initial 'that can't be true' sort of reaction. But what you will find is that there are requirements and then there are expectations.

So your first big task is to define what the **requirements** are. And to do this you need to work with your partners and work with your consultants. Benchmark: Crucial for our experience in being able to understand what the true requirements are is to benchmark comparable industry practices.

We have also been able to find English translations of Japanese regulatory documents. There are websites that exist that have such things. While you don't have the chance to speak formally with MHLW or other regulatory agencies in Japan, you do have the opportunity to speak informally. To do this you need an entree. You need your partner company in Japan to go to the agency and initiate an informal discussion, and then you can be invited in to tell this person about your company. This is all to do with the trust issues, with the familiarity issues, for Japan that are very important.

So once you have decided what your requirements are and what are the expectations or 'nice to haves,' for requirements you had better try and follow them if you can. Consider harmonization with other regional requirements. If Japan requirements cause you to do twice the stability work that other countries do, you might have to see if it is feasible. You might have to prioritize and your Japanese colleagues can help you with that. If you have resource constraints you might not be able to do everything. But if you don't do everything, you need to make sure that decision is clearly understood up front.

For **expectations**, again consider the cost of compliance with the risk of non-compliance. And in this case, some of us may not understand the powerful position that the Japanese clinical trial director holds in terms of not only implementing the clinical trial but in your own good name as a company. So if your Japanese colleagues at your partner company indicate that there is a big problem here with not meeting an expectation, it could be because of this comfort level factor – that your good name is very fragile in Japan and you really need to be sensitive to your colleagues when they tell you that, 'oh we really have to do this,' because they might be absolutely right.

Also I just wanted to say document everything again. It is especially important with working with development in Japan because, as I said, you do not have the opportunity to update the agency regularly, so you need a good record of all your changes.

Examples of issues that are very common: Repetitive testing…. Japan regulators customarily have expected that you will do three replicates of your test even if you have a test that has one replicate. You will do three of those tests on each stability time point. Seems a little counterintuitive because the guidance tells us we should be using our method validation information to decide how many replicates are required. However, that has nothing to do with it for Japan. It is an expectation but it is a very strong one. I am not saying that you need to do it but you do need to be aware that you are going to have a conversation about this.

Full specification testing at least yearly is another thing that is often asked for and again, something that will require a lot of negotiation if you don't want to do all of your specification testing every year. You will need to have a good justification for why.

I think that probably one of the biggest issues for Japan regulators and Japan industry as a whole is that they don't really understand biotech as well as other jurisdictions might. And so they do tend to adopt a lot of small molecule guidance inappropriately. A lot of what you need to do involves mentoring and educating and trying to develop a relationship with regulators informally so that you can explain your company and explain your goals.

Related issues: Specifications. This is something that is an issue because in Japan, your specs are usually set on your three conformance or confirmation lots, and your supporting data is not considered to be that relevant. So you might end up with tighter specs in Japan. Use of JP reagents and test methods, cosmetic appearance and particulates are all issues. The cosmetic appearance is not a quality issue. It is just that the clinical trial directors expect to see commercial-looking material at Phase 1. They want to see blister packs if you have got pills, and so that is going to be an issue.

Overall Benefits

I'll just say a little bit about a couple of things that we probably want to touch on in our talks. **Out-of-trend investigations**: There is really no consensus about this that I am aware of. There is no real guidance out there for us. There are some papers that have been developed by a PhRMA group. What level of monitoring is appropriate for biotech products? How do you identify out of trend and what assays are most significant for monitoring?...In this case, stats are not well established or widely used for this sort of activity. My point about the design space and the building in quality is that the better that you understand your product and the better you understand your assays, the better you'll be able to deal with out-of-trend issues, which you need to deal with.

Same with non-conformist resolution. The bottom line is that you will, by understanding your product better, by having more robust programs that started earlier in development, be able to address non-conformances in a more comprehensive way. And then of course, improving specifications: Specifications and stability are tied. You cannot get to commercialization without having all the tests be relevant and non-value added tests gone. You need to have that happen. The whole point of this is to reduce the risk of accepting unsuitable product or rejecting acceptable product.

So in **conclusion**, a comprehensive stability program includes the program and the strategy – what you have done, how and why you have done it, and document your rationale. There is a whole list of benefits which I've talked about.

A comprehensive stability strategy requires more early development work and organization. So it needs advocacy by those of us in technical fields. The work must be value-added and you have to keep the goals in mind. You partner with other functions. Sometimes you need service level agreements to make sure that things happen. You definitely need a good project plan and good program management. The benefit you will have is that there will be increased visibility to commercialization goals and improved teamwork.

Documentation: Again document everything, including your project plans, decisions, rationales, product information and also contingency plans.

It will provide a roadmap to commercialization, fewer surprises, a more robust submission, an increased ability to focus on phase-appropriate studies, incorporate global requirements as you need to, ensure that the tests are appropriate and improve your specs, address nonconformances, and evaluate out of trends and apply for post-approval assay relief as appropriate.

"Let's assume that you have gone through your clinical development profile. You have done all these studies. You have shown that some of these assays are not stability indicating. At what point post marketing," Advant queried, "can you actually take them off? Is there precedence from the industry that they have been able to take them off?" A related question, he said, is "can you take off some time points if you are doing, you know, zero, three, six, nine, twelve? Can you do minimal time points if you have shown that the product is stable?"

Genentech's Nashabeh responded "basically yes to both." Typically the assays are removed by Genentech prior to licensure, he explained. Based on all the cumulative data from the clinical phase, "we will adjust the number of assays and type of assays that are on the commercial stability protocol going forward."

Addressing the follow-up question on reduced sampling, he said that "it is not unusual that for commercial [product], for the annual commitment, that you may not have the same time points, the same frequency of testing as you would have in early development. Whether this occurs right at licensure or later on – that you will actually reduce the time points – is a matter of how much confidence you have in the prior data. That is how we have kind of done it over the years."

➢ The post-marketing session participants discussed further the implications involved in making changes in the testing regimen as the analytical technology evolves.

Reiterating the agency's general interest in encouraging sponsors "to come up with better, more reliable, more precise assays," CDER official Shacter emphasized that "the trick is to be able to demonstrate the continuity."

The question, she said, is "are you really measuring the same parameter with both assays, or are you measuring different specific molecular characteristics? And if you are not, how can you link them, how can you go between them so that you have, again, continuity of your trending data – so that you will know that all of a sudden if you have a change in the stability data, was it a true change in the molecule or is it because you introduced a new assay that either has greater sensitivity or is measuring actually a different element of the molecule?"

Shacter recommended that sponsors "bring that change to the agency and discuss it with us so that we can come up with some good solution. Certainly make a proposal, teach us about what you are looking at, what the change is, and then we will work with you to institute the change."

➢ PharmaNet Consulting Executive Director William Egan, who recently left the agency after an extended career in CBER review management, followed Shacter's comment by again emphasizing the need for retention samples.

"Having the retention samples to do the other comparisons is extraordinarily helpful because the material that has gone into the clinic and the material that has been going into people has not changed, only our ability to characterize it – being able to detect…the things that are actually there, whereas previously you could not. But the matter itself has not changed, only our perception of it. I think one has to keep that in mind."

Define Purpose and Goals Of Stability Program

Egan and Shacter gave presentations at the beginning of the post-marketing stability session that helped inform the ensuing discussions.

Egan stressed the need for understanding and carefully defining the purpose and goals of the stability program, while Shacter discussed the implications for changes to expiry and to the product, process, or analytical regimen. Shacter also provided case studies of stability problems that have come to the attention of the Office of Biotechnology Products regarding the products it regulates (*see box on pp. 23-28*).

➢ Egan emphasized that understanding the purposes and goals of ongoing stability studies is necessary to determine how those studies should be set up in terms of what is evaluated and the number/spacing of the sample data points and the algorithms involved. "So I think it is actually fairly important to first answer the question, why are we doing these studies? What do we hope to get out of them? And then, how can that goal best be accomplished?"

The design of the protocol, in turn, will be dependent on those purposes and goals. The questions at issue, Egan pointed out, include: ● "Are the data from the annual studies being used, pooled, to better define the shelf life at some period?" and ● "How does one define 'out of specification' for any ongoing study? Does it refer to single data points or collective measures of those data points? And again, is that context dependent – dependent on the study that was set up?"

In making such OOS determinations, Egan continued, "certainly, if you have an ongoing stability study of three points, one at the beginning, one at the end, and one in the middle, and the one in the middle is out of specification, there is not much choice about how to

deal with that. If it is an ongoing stability study and there are 50 data points and one of them is out, and the rate is very much the same as what had been observed historically, then that is not an issue. So it depends on exactly the study that was set up."

One question is whether the protocol can be defined to permit additional testing as warranted if a few points are starting to appear out of the expected range. "That protocol," Egan explained, "has to be set up so one is not testing into compliance, but rather trying to better define a particular measure – namely what is the quality of that product – and then answer the question about the consequences of being out of spec on other manufactured lots if that one lot is taken as representative of everything else that is being manufactured."

Egan summarized that ongoing stability testing programs "should be very well constructed to meet their intended goals, and those goals should be defined prospectively. And to the extent possible, I think we need to define how we are going to deal with various situations that might occur during the ongoing stability study – for example, a data point or points that fall below a certain potency limit. How are these going to be handled?" He stressed that "these stability studies are resource intensive, particularly if they are done well, so I think they should be thought through as to their purposes."

Looking to the future, Egan suggested that "what is needed is a continued dialogue on the purpose of annual stability studies and how to best achieve that purpose or purposes."

➢ In her presentation, Shacter cautioned manufacturers not to think of the stability protocol as locked in stone and not needing amendment.

"A point to the wise and that is that change begets change," she advised. "Your stability protocol may have been very well designed for the original manufacturing process that you were using, but when you make a change to the process you could actually impact the relevance of the stability protocol. So you may need to modify the stability protocol to stay in concert with how you have changed the process."

For example, where formerly a firm did not need to be continuing to look at deamidation of the product, if a step is added that increases pH in the process, "you may need to go back and reassess that for stability," Shacter explained. "And similarly, if for some reason you have an extra high concentration of your protein during the process you may need to do additional aggregates testing."

The CDER biotech official further cautioned that changes in the manufacture of the drug substance will generally necessitate putting drug product lots made from that bulk on stability.

Shacter also provided the participants with some guidelines regarding making changes to stability tests: "If you are going to change a test, you should demonstrate, needless to say, that the test is either equal to, or better than, the original test that you were using. And if you want to remove a test, then you need to demonstrate that lack of utility of the test that you have been performing."

➢ In line with the forum discussions, Shacter recognized the degree of concern among manufacturers over the number of stability tests that need to be done prior to and after licensure.

"We have talked a lot about the need to test orthogonal methods or any method that shows a change in stability, and I know the concern among manufacturers is that you will always have to do those tests…into perpetuity." However, she stressed, "that is not the point. I can guarantee you that we will consider the data showing the lack of utility of the stability test, so that if it is really not relevant, if it is not teaching us anything about either a critical quality parameter of the product or one that simply isn't variant even though it is critical, we will talk with you about taking that out of the stability protocol."

If a particular reviewer proves recalcitrant, Shacter urged participants to refer to the meeting concensus on this point, because "we don't want to have a burdensome number of stability tests ongoing." On the other hand, she stressed, "it is so important to have that information before licensure so that we can understand the product and know the meaning of any changes that we see."

➢ Another important point made by Shacter relevant to changing a stability test – either introducing a new one or replacing an old one – is to do head-to-head comparisons.

"Don't rely on historical information from the former test," she warned, noting that "many people do try to rely on historical data versus doing side-by-side comparisons of the test." She urged sponsors to "accumulate as much comparative data as possible so that we know we have a basis for assessing the change in the test, and then submit the changes as a prior approval supplement to the agency." ♦♦

CDER's Shacter on Biotech Product Post-Marketing Stability Issues

The following talk on post-marketing stability issues was given by Division of Therapeutic Proteins Biochemistry Lab Chief Emily Shacter at the WCBP CMC Strategy Forum in late July. Shacter discussed the implication of manufacturing and methology changes on the stability program and cited stability-related problems encountered among the products regulated by CDER's Office of Biotechnology Products.

What I am going to focus on is what happens when you have a change in the manufacturing process, either intentional or unintentional, and what you should be doing about that. This would be a change to the process, to the product or to any assays that you are using as part of your stability protocol. The main goal of the stability studies in this case is to demonstrate comparability and continuity of both the product and your assessment of stability of that product.

I will also talk about some stability problems that have become apparent during licensure and focusing somewhat on managing and understanding OOS results, out of specification results, and some findings of inadequate stability testing. At least half of my talk will be giving you some case reports of various things that we have seen at the FDA that have happened in the course of stability testing of a product.

So just a little bit on **guidance** to sort of put everything in focus: It seems a little late but on the other hand some of these things actually have not been said so far in the meeting. The setting of specifications for product release needs to ensure that the product continues to have the safety, purity and potency reflected in product labeling. The emphasis for the purposes of this study is that it has to continue to have those qualities. That is the purpose of stability testing. The other most important point for us to remember is that a lot of our statutory ability to even regulate stability comes from the Federal Food, Drug and Cosmetic Act, Section 502(a) where it says that 'a biological drug product is deemed misbranded if its labeling is false or misleading in any particular.' So if your product has changed to the degree that the label is no longer reflective of the product, then the product is misbranded and you will have a compliance issue associated with that.

Advice On Making Changes

One of the common questions that comes up is **'how do you extend your expiration dating period?'** Now we are talking post-licensure, and expiration dating is obviously a critical parameter, whereas yesterday when we talked about our early stages of clinical studies, 'well what is an expiration date, really?' But now we are talking about true expiration dates both for the drug substance and the drug product.

Really, all you need to do in order to extend an expiration date is to modify the stability protocol to add the longer time points, provide the updated stability data and any trend analysis that you have on those data to the agency, and if you have an approved stability protocol and you have data that meet the acceptance criteria showing that extension of the expiration date is in fact an appropriate thing to do, then you can extend the expiration date in an annual report. This obviously assumes that no new trends have been observed over the longer time course of looking at the stability of the product.

If you make a **change to the manufacturing process or product**, you should be putting those new lots on stability. The number of lots you need to put on stability depends on the nature of the change that you have made in the process. A significant change in the process might require something like two or three lots, whereas a relatively minor change in the process which is not expected to have a significant impact on product quality, probably one lot will do. We gauge those on a case-by-case basis depending on what we think the risk is to the stability of the product.

You should perform accelerated and or stress testing along the lines that we have talked about already during the meeting. You should do trend analysis across the lots to see if the manufacturing change has had any impact on what you know about stability of the product. You should analyze the data – and this is the most important point that I would like to make – using prospectively defined, statistically based acceptance criteria to assess the comparability to the previous process. It is not sufficient to take the accelerated stability data for example, to look at the lines compared to the old material and say, 'yes, they look similar to me.' That won't be acceptable. The other

point I would like to make is that you should really be assessing again, as we have discussed, the rates of decay and not just end points over the time course of the stability testing.

After you have done all this you should submit a supplement to the agency. The level at which that supplement needs to be, whether it is an annual report, a CBE-30, or a prior approval supplement is determined based on guidance and CFR 601.12. One of the easiest routes is if you have a comparability protocol, then you can most likely submit the new information in the annual report.

A point to the wise: Change begets change. Your stability protocol may have been very well designed for the original process that you were using, but when you make a change to the process, you could actually impact the relevance of the stability protocol. So you may need to modify the stability protocol to stay in concert with how you have changed the process. Some simple examples are if you added a step where you have an increased pH in the process where you formerly did not need to be continuing to look at deamidation of the product, you may need to go back and reassess that for stability. Similarly, if you for some reason have an extra high concentration of your protein during the process you may need to do additional aggregates testing. Be sure that your stability protocol will continue to support any change that you make in the process.

A simple point: If you make a change to the manufacture of the drug substance, you also need to put drug product lots from those new drug substance lots on stability. That is a general rule.

Now if you want to make a **change to your stability tests**, a couple of guidelines: If you are going to change a test, you should demonstrate, needless to say, that the test is either equal to or better than the original test that you were using. If you want to remove a test, then you need to demonstrate the lack of utility of the test that you have been performing.

On this point, I know that there is a lot of concern among manufacturers over the number of stability tests that need to be done prior to and after licensure. We have talked a lot about the need to test orthogonal methods and lots of methods or any method that shows a change in stability. I know the concern among manufacturers is that you will always have to do those tests…into perpetuity. That is not the point. I can guarantee you that we will consider the data showing the lack of utility of the stability test so that if it is really not relevant, if it's not teaching us anything about either a critical quality parameter of the product or one that simply is not variant even though it is critical, we will talk with you about taking that out of the stability protocol. You can tell your reviewer since I know that there are some differences, depending on what reviewer you get, but just refer to this meeting because we do do that. We don't want to have a burdensome number of stability tests ongoing. It is so important to have that information before licensure so that we can understand the product and know the meaning of any changes that we see.

If you are going to change a stability test, introduce a new one or replace an old one, you should do head-to-head comparisons. Don't rely on historical information from the former test. Although that seems obvious probably to most of you, you would be surprised at how many people do try to rely on historical data versus doing side-by- side comparisons of the test. You should accumulate as much comparative data as possible so that we know we have a basis for assessing the change in the test. Then submit the change as a prior approval supplement to the agency.

I think an interesting question that comes up is 'how do you set comparable specifications for different types of end points?' For example, if you are looking at charge variants and you are going to replace isoelectric focusing, which is an assay with an awful lot of variability, with HPLC, which might give you more reliability, how are you going to compare the end points from those two tests since they are really very different?

There is not an *a priori* answer to that but it is obviously an interesting question for everybody to be thinking about. The important element here is to make sure that you have continuity. It is to make sure that the tests are really measuring the same thing, so that you know your changeover from one test to the next is still measuring the same parameter, and that you have continuity when you are doing trend analysis of your lots over time and over the course of many years. You want to be able to refer back to your older data.

Examples of The Stability Problems FDA Has Seen

The rest of the talk will be giving…case reports of some incidents that we have seen that relate to stability in the many products that come through the Office of Biotechnology Products. I didn't get any examples from the Office of New Drug Chemistry, but I assume that the stories would be rather similar.

The first one is a case [in which] the sponsor wanted to **change from a lyophilized to a liquid formulation**. In making that change they needed to now start testing the vials in the inverted as well as the upright configuration. What they observed was an increase in protein degradation over time. The reason for this was that now that the fluid was touching the stoppers, there was leaching of a metal from the rubber stoppers and it was activating a metalloproteinase that was present in the product. That metalloproteinase was probably always present in the product, but because there wasn't enough metal there to activate it, the changes in degradation were not seen until they went to this change in formulation. In this case, the resolution of the problem was to add a chelator to the formulation and that was acceptable.

Probably one of the most famous recent cases of a serious problem that we have had with a product is the story of ***Eprex*** and Pure Red Cell Aplasia. And I'm sure that most of you know the story but basically there was a **change in formulation** for this erythropoietin product – which is marketed in Europe, not in the United States – from a formulation containing human serum albumin to a formulation containing Tween. They used the same container closure system, which was a pre-filled syringe that had uncoated rubber stoppers. After the course of this change, basically what was found was that patients were coming down with a severe and life threatening disease, which as it turned out – at least the current hypothesis on it, with some good data but nobody's absolutely sure still what the cause has been – there was a leaching of organics from the rubber stoppers causing a change in probably immunogenicity of the product.

We don't usually talk about the Eprex story as having been a stability story, but the truth is that there was an increase in leaching of these organics from the rubber stoppers over time, so that at early time points the levels that were measured by HPLC were relatively low and they went up over time. There may be a correlation between the patients that actually experienced PRCA who were receiving old lots of material versus younger lots of material.

One point to make here is that there was no observable change in the drug product. The hypothesis is that the leachates that were injected into folks were actually serving as adjuvants and heightening the immunogenic response to the protein. Again this is hypothesis – it's not known, there is some data to support it – but the point is that actually the testing of the drug product wasn't detecting any change there.

The point of these two stories is that you need to revalidate the system compatibility and the stability of your drug product and drug substance if you change your formulation or the container closure system. What I think we learned from both stories is that you need to be able to examine changes both in the protein and in the impurities profile, because both of those – pretty serious changes to the product – resulted from impurities and not from product-related changes.

Another story where a sponsor wanted to go to a **low dose version of a lyophilized product**. This is a product that was very stable at 15 to 30° C, had a long stability history, and in going to the smaller configuration the vial was smaller and the content was smaller, but they used the same container closure system. But what they found was that this smaller configuration was not as stable to store it at 15 to 30°, basically controlled room temperature, as the larger vialed configuration. It turned out that they needed to be storing this lyophilized product at 2 to 8° C instead of room temperature.

We don't understand the reason for this – possibly the ratio of the container closure system to the amount of protein in the vial. But the point is that you wouldn't necessarily expect it, but you can have a change to a lyophilized product just by changing the vial configuration.

Another observation was a series of data that were showing that there was a 20% **loss in activity** of a product – this was a cream in this case – over the shelf life. These were data that were discovered through post-marketing commitments to look at the ongoing stability of the product. What happened was because there was such a dramatic change in the activity of the product, it went out of spec while the product was on the market. It would have been recalled if there was any product left, but I believe that in this case, there wasn't any left. There were several lots that went out of spec. [They] were already used up by the time of the 12 month shelf life. So there wasn't any actual recall.

The resolution to the problem was, I think, a little bit creative, and that was that the sponsor added an in-process limit. So they did not change the release specification. They instituted an in house in-process limit that required that if the product, on release, was lower than a certain level, that lot could go out to market but those lots would have to be placed on stability. If the lot fell out of specification while the product was on the market, it would have to be recalled by the sponsor.

The next…is a case where the **pre-market stability data were really inadequate**. The sponsor extrapolated from real-time data to set an expiration date which was set wrong. This is a case where the product was actually gaining activity over time. Also to place as a reminder that products also gain activity over time, and it is not just a matter of losing activity, but also gains of activity.

This product gained activity in half of the expected time. So they had had a 24 month shelf life but within 12 months the product was out of specification. A re-analysis of the data following the guidelines in Q1E, showed that actually, the expiration date was not set appropriately and that the product had a shorter than expected shelf life at room temperature. The problem was probably aggregation in this case. I'm not sure that that was fully established, but the resolution to the problem was to change the shipping temperature and ultimately the storage temperature to 4° centigrade.

I think in this case, we don't like to admit this, but I think the FDA made a mistake in allowing the extrapolation that was made. So this really shouldn't have happened, and, in retrospect, I think that we will definitely be much more careful in terms of using the appropriate statistical criteria in setting a shelf life based on extrapolated data. We certainly don't like to do that at all. As a general rule we will extend shelf lives an incremental amount. For example, if you have 15 months of data, a shelf life of 18 months, if you know the decay rates, is a reasonable extension. But to go, for example, from 12 to 24 months is something that at least in DTP [Division of Therapeutic Proteins] we never do, and I am sure that DMA [Division of Monoclonal Antibodies] doesn't do that either. So lessons learned even for the FDA.

Now I have two examples of cases where products went out of specification for potency, largely due to **poor potency assays**. In this first case the potency assay had a very high degree of variability. No surprise, a number of lots were tested out of specification for potency while the product was on the market. Unfortunately in this case, this was a relatively poorly characterized molecule, so there were not good physical chemical tests to be able to tell us or the sponsor whether it was truly missing its potency specification or whether it was really an error due to a lousy potency assay that had an awful lot of variability. So there were multiple product recalls in this case. The rule of thumb coming out of this is that the looser your bioassay, the more chance you have of getting an out-of-specification result. We would encourage you, and I'm sure you want to have the same, to have the tightest assay limits that you can establish.

I know a lot of people talk about how widely variable potency assays are. I have to say that in my experience at the FDA, that is really not necessarily true. Even for assays that are around that are proliferation-based assays, so complex cell-based assays, it is not the norm to have the assay variability be in the range of two to three fold. Now some of the control comes from using an internal reference standard. But that is part of having to deal with the complexity of these assays. But I actually personally don't buy the argument that potency assays are inherently variable.

The second example of an unreliable potency assay resulting in out of specification results: This again was an enzyme that appeared to gain activity at a higher than expected rate during storage. There were both multiple product recalls and failures to release the product. In this particular case, there were two concurrent problems happening. One was that the potency assay was very inaccurate. It was actually giving them false high results, just by virtue of the design of the assay. It was a really poor choice of potency assay. But the other issue that was happening was that the product was unstable. It actually truly was unstable. It was a conjugate and the conjugate was falling apart, and as it fell apart that also changed the activity. So there were two concurrent things going on which made it very difficult to be able to predict what was going to be happening.

The resolution to this problem was to both improve the potency assay – to completely change it – so not just find better ways to tighten up this particular assay but to ditch the assay entirely. It was not destined to be a reliable assay in the first place. Then also to improve the molecular stability of the product. Obviously what you want to be doing is minimizing the variability of the assay and the potential for instability in your product.

One other point to make in this case is that since this was a conjugated product, obviously you need to test for the stability of the conjugate. Although that seems like an incredibly obvious thing, surprisingly that was not done sufficiently before this product went to market.

Ad page.

Now a case where we had an **invalid process validation** leading to products going out of specification while on the market: What happened here was that the manufacturer had in theory validated through small scale studies that a pH acidification step could go down as low as pH 2 and the product would be fine, there would not be any stability issues. But in fact what was observed was that marketed lots in which the material had been exposed to pH 2.2 actually were going out of spec for aggregates at pretty early time points – in fact early enough that it's probable that some of these lots should have been detected upon release, but because the aggregates assay was not terribly reliable, they actually were not testing it even on release. As a result of this, product lots were recalled and the manufacturer, to resolve the problem, basically raised the lower limit of the pH to which the protein could be exposed for viral inactivation studies.

Another case where there was a product that had multiple different product strengths. Issues were coming up for the **low strength vials that were reconstituted** in bactericidal water for injection. What was found was that the low dose versions were going out of spec for protein. In fact they were losing 20% and more of the protein in the vial. This is not something that we see very often. Most often, parameters such as pH and protein are pretty rock solid. So this was an impressive loss of protein, possibly due to aggregates in the protein coming out of solution, and it might have resulted from an interaction with the benzyl alcohol. The point was that the manufacturer hadn't validated that this product could be reconstituted and be WFI at all. So there were multiple product recalls.

Again, the resolution of this problem was for the sponsor to validate the use of WFI for reconstitution of the low dose vials. This was submitted to the agency as a prior approval supplement. The sponsor also needed to increase the protein release specification by 20%. So instead of being something like 90% to 125% of the labeled amount, it actually increased to 110% to 125%.

Now the sponsor came in and they said, 'Well, actually, we would like to loosen the specification for aggregates, because when this product gets reconstituted [in] WFI then the product goes out of spec for aggregates. But that is what is going to happen so can't we just change the specification for the allowable amount of aggregates?' Needless to say the answer was 'no.' If the specification is set wrong in the first place, then you can't retroactively come back and say, 'Well gosh it was always that way, therefore we want to change the spec back to what we should have known at the time.' That is not allowed.

We also do identify **stability-related issues on inspection**. I'll give you a couple examples of those. In one case the sponsor was testing the stability of a **process intermediate,** but the test samples were not being stored in the same configuration and the same conditions as the product intermediate was being stored. This was a 483 citation because, again it seems fairly obvious, the test samples need to be stored under the same conditions as the product itself.

In another case there wasn't any stability testing of a **protein excipient** that was used in the final drug product, even though it was known that this excipient would affect the PK and PD of the active ingredient. In fact that is part of why the excipient was there. The sponsor should have been testing the stability both of this critical excipient ingredient as well as the active ingredient. So that was a 483 citation.

We have also had any number of 483 citations come in from the field because of inadequate stability indicating assays in a stability protocol.

Now there is one 483 citation which we might have some reasonable disagreement on, and that is that we have seen some 483 citations coming to us for failure of a sponsor to do **endotoxin testing** as part of the stability protocol. The scientific rationale for requiring this testing is unclear to us. So there is a little bit of misunderstanding I think in the field about why you would or would not have to have that in there. If this happens to you while you have the field inspecting your facility, ask the inspector to please call the product review team so that we can hash this out with them because this is a 483 citation that does not need to happen in our perspective.

So some cheap **advice:** You should be establishing and optimizing meaningful and quantitative stability-indicating assays. You should plan for stability issues when making any change to the process or to the product. And you should institute appropriate stability testing as part of a process change. Don't operate on assumptions, because we at the FDA don't....As we like to say, 'Just show us the data.' We hope that we will exercise our own scientific acumen to negotiate with you about what the appropriate response should be.