MINUTES OF THE MEETING OF THE PQRI PSD MASS BALANCE WORKING GROUP ON 20 JULY 2004

I. PARTICIPANTS

Terry Tougas (Boehringer Ingelheim), Chair
Dave Christopher (Schering-Plough)
Paul Curry (USP)
Craig Dunbar (Alkermes)
Ken Furnkranz (FDA)
Zoë Heaton (Aventis)
Rick Lostritto (FDA)
Lana Lyapustina (IPAC-RS)
Jolyon Mitchell (Trudell Medical)
Brian Rogers (FDA)
Helen Strickland (GlaxoSmithKline)
Bruce Wyka (Schering-Plough)

II. OPENING

Dr. Tougas welcomed the participants and opened the meeting. Dr. Lyapustina reminded the participants that their discussion is subject to the anti-trust guidelines applicable in the United States and European Union, and that nothing discussed at this meeting may be intended to restrict trade or individual decision-making of any company; she further instructed the participants to avoid discussion of competitively sensitive subjects, such as confidential marketing, sales, and pricing information.

The host, Dr. Lostritto, also welcomed the Working Group members and expressed his appreciation for their time and work being contributed to this effort. Dr. Lostritto offered a few reflections on his role as a new Team Leader at the CDER Pulmonary Division. He stated that he and his Team, as well as the Division and the Center, stand behind innovation. It is this changed climate that attracted Dr. Lostritto to the present position, he said. Dr. Lostritto explained that having worked in the industry for some years, he understands manufacturing risks and realities, and is a proponent of a science and data-based review system, with appropriate regulatory risks. He stressed the importance of data in bringing about change.

The participants agreed on the following objectives for the meeting:

(i) to review the draft Mass Balance recommendations and submitted revisions;

(ii) to discuss statistical comparisons of variability of mass balance vs. emitted dose measurement;

(iii) to discuss mechanism for implementing mass balance as a product specification vs. system suitability vs. run qualification;

(iv) to draft consensus recommendations;

(v) to discuss leadership of the Mass Balance Working Group in light of Dr. Tougas' new role as Chair of DPTC;

(vi) to discuss future plans ("a map for going forward") for the Working Group; and

(vii) to agree on next steps.

III. DISCUSSION

Testing for Defects vs Testing for Quality

Dr. Tougas invited the Working Group to consider the philosophical objective of the end-product testing in general and the mass balance testing in particular, as one aspect of end-product testing. The Working Group discussion is summarized below.

Pharmaceutical quality control testing on the finished product should by its nature be approached as confirming the quality of the batch. For this reason, end-product testing is limited in the number of samples tested. Not only are large sample sizes impractical, they are also inappropriate for confirmatory testing (especially when non-parametric tests are used because the chance to fail then depends on the sample size regardless of batch quality). Also, testing for defects has been moved back into the manufacturing process in the testing of the components.

By contrast, testing for defects, of which testing per military standards (MIL-STD) is one type, requires large numbers of samples, whose size depends on the assumed percentage of allowable defects. The sample size of the quality confirming testing, on the other hand, should be sufficient to estimate batch parameters such as the mean and variability, and for this purpose much smaller samples are appropriate. Quality testing allows sponsors to make an informed decision about disposition of the batch, but it cannot be used to detect defects.

Testing for defects and end product testing to confirm quality have different objectives, ask different questions, and therefore should use different test philosophies and test structures. For instance, because of the large numbers per sample involved, defect tests must be simple, lending themselves to a fast and easy implementation. In the case of inhalers, testing whether a valve sprays or not, could be done quickly, whereas assessing DCU takes significantly more time and effort; and similarly, CI testing is very time- and labor-intensive. Some CMC tests are indeed testing for defects, such as stress testing, gross leak testing, weight checks, and spray-testing, which could be done on-line, 100%, focus on one unit at a time, and use a different statistical procedure than a finished product QC test would. The end-product testing, by contrast, assesses the batch, not unit, characteristics. Hence end-product testing should typically focus on mean and variability (of the batch, not of the unit). Other industries, such as semiconductor industry, use parametric batch quality testing rather than MIL-STD type defect testing in this context. The working assumption is that batch properties are well understood, controlled during manufacture, and deviations from the expected behavior are rare.

FDA participants noted that although they are all for the modern way of thinking, currently, reviewers may not have a very good picture of how the batch manufacture is controlled, and hence may tend to adhere to old approaches. Historical inertia may play a role as well. In addition, FDA has to deal with adverse event reports and complaints from patients who get that one-in-a-million super-potent or under-potent inhaler. The Working Group noted that if 100% testing is the goal, other methods and tests could be used, e.g., weight check. In the end, however, it is impossible to guarantee a 100% error-free batch. Even heart valves fail, albeit with a very low frequency. This is unavoidable, as the only alternative would be having no product sold at all.

The old regulatory paradigm focused on looking for "the one bad apple," and when one was found it was assumed that the entire batch was bad. By contrast, the new paradigm acknowledges that all characteristics and measurements are distributed over some range, and a single result outside prescribed limits could still be within the natural range of variability. Therefore the 21st century regulatory paradigm is about building quality in early, and only confirming it through end-product testing. Batch quality is *ensured* by appropriately designed and controlled manufacturing processes, including on-line/at-line testing strategies. The end-product testing is used to confirm the expectation that the batch quality is appropriate. It is this philosophy and the associated new view on objectives of testing that underpin PAT, and not simply a particular analytical method, e.g., near-IR. One cannot test for low-frequency defects at the end-product testing using complicated analytical procedures.

One consequence of this approach to quality is that zero tolerance criteria associated with end product testing are ineffective either at assuring batch quality or at detecting low frequency critical defects, and serve only to increase producer risk (i.e., rejecting a good batch).

Mass Balance measurements should not be viewed in isolation from the other tests and controls conducted on the batch. For example, mass balance is one way to capture emitted dose, which is also controlled through a specifically designed and more efficient test for delivered dose uniformity, as well as through control of batch assay, orifice dimensions, shot weight, etc.

The participants also discussed whether statistical theory based on normal distribution could be relied on when considering MB or DDU. Mr. Christopher explained that by the time a company goes through development, the distribution is well-characterized, and even though it may not be perfectly normal, the normal statistical theory is robust enough to deal with all practical situations. Distributions may deviate from normality with some statistical significance, but that deviation would be insignificant practically and should not affect decision making. Mr. Christopher further stated that after 25 years of applying statistical theory based on normal distributions he is convinced of its robustness for making correct decisions. He clarified that sometimes observed distributions may appear curtailed on one side (e.g., due to approaching limit of detection or quantitation), but this is due to the physical limitations of the method while the distribution itself is normal enough for practical applications of theory. In answer to Dr. Lostritto's question, participants clarified that specifically in the case of mass balance results, the distribution may be centered at lower than 100% (of the target emitted dose), but the shape of it would still be close to normal – the shift does not affect the shape.

Mass Balance Recommendations and Flowchart

Dr. Dunbar and Dr. Lostritto each presented revised draft recommendations with revised flowcharts, which outline recommended actions upon observing a mass balance "failure." During the Working Group discussion, the following points were made:

- All types of failures are recorded in the lab notebook, along with any investigations undertaken and their results. In particular, all assignable causes are captured in the lab record.
- Companies' own office of compliance would not allow numerous retesting.

- If a method-related cause is found, up to three cycles of retesting should be allowed, and all results from the previous run(s), both MB and APSD, should be invalidated, because at this point, the product has not been tested yet, but only the system. Failing a batch when the problem is caused by the method will not solve the problem.
- System suitability tests (e.g., assessing precision, accuracy and resolution) use standards, and in this respect they differ from the mass balance determination. Pure system suitability tests never lead to an OOS investigation or batch failure. On the other hand, a mass balance determination from CI testing depends on suitability of the testing equipment at the time of the run, as well as on the sample being tested. It therefore cannot be viewed strictly as a system suitability test but neither can it be viewed as strictly a product specification.
- The procedure for retesting is normally included partially in the method description, and partially in company's own SOPs.
- Testing for emitted dose could be one way to confirm that the canister in question delivers the intended dose. However, this test would have to be of a different structure than the standard delivered dose uniformity test, as it would have to be performed on the same canister as the CI test (multidose products), use the same number of actuations as the CI test, at the same life stage as the CI test, and it may need to use different limits than the standard DCU test. This "single-canister emitted dose" test would need to be separately validated, which was viewed as a drawback by the industry. A different scheme would need to be applied to pre-metered or single dose type products. Several participants also questioned the reason for focusing on the one canister with failing MB if the purpose is to evaluate the batch, especially since a CI test would need to be re-run any way. The aim of evaluating the batch would be better served by testing a new canister. (The participants discussed, in particular, a hypothetical scenario where the canister with the failing MB in fact delivers 76% label claim consistently). The large number of doses needed for DCU testing on a single canister was discussed as an additional complication, as the canister may not have enough doses left for all such retesting. In response to a question whether chances are higher to pass a "single-canister emitted dose" or the repeat CI test, Ms. Strickland explained that this depends on the relative contributions to the variability from the product, CI test and DCU test, at the given number of actuations and testing protocol, all of which would be product and method specific.
- The limits of 85-115%, 80-120%, 75-125%, 82-128%, were selected arbitrarily.
- The origin of 85-115% may be traced to the similar limits on the mean for DCU determinations, however, there the mean refers to an average of 10 or more determinations. It may not be possible to meet the 85-115% limits based on a single dose in a CI test. The language in the final CMC nasal guidance may be appropriate to address this situation: "*If the procedure is based on a single actuation determination, then the range can be broadened to reflect the limits allowed for an individual actuation.*"[1]

[1] http://www.fda.gov/cder/guidance/4234fnl.pdf, p. 25 of 49.

- The specifications for stage groupings are data-driven, and usually the upper to lower limit ratio is 2:1. These limits, however, do not add up to the recommended mass balance limits. On the other hand, unlike stage grouping specifications, the mass balance captures both sizable and non-sizable components and therefore having a separate set of limits for MB seems appropriate.

- When setting limits, FDA currently is taking into account the therapeutic indication, so that ranges for steroids, for example, may be more critical than ranges for beta agonists.

- Recentering the limits to, e.g., 80-110% may solve many problems. The 5% offset is consistent with USP's allowance of up to 5% for CI wall losses. For some products, however, the test may lead to high MB rather than low - for example, in suspension and solution MDIs at the end of container life, or for DPIs that are emptied by the analyst manually via opening the plastic container, which may lead to electrostatic charging. Issues like these are product-specific, therefore specific limits should be discussed with the applicant. Nevertheless, some default limits would be appropriate to have as general guidance, and ±15% seems reasonable.

- "Outliers" may be aberrant results due to some testing errors, or observations from the tails of the distribution (which nevertheless is a good distribution), or indications of additional distributions being mixed in with the main underlying distribution. The applied tests and limits should be based on the understanding of the behavior of the batch and should allow for expanded testing to understand such behavior without penalty.

Statistical Comparison of Mass Balance and Dose Uniformity Limits

Ms. Strickland presented a statistical analysis of the various limits imposed on canisters through DCU and MB tests (see Exhibit A). The analysis shows that the ±15% MB requirements are more stringent than any of the DCU requirements. For example, under one of the used scenario assumptions, meeting DCU's 75-125% limits requires a total variance of less than $69\%^2$, meeting the DCU's 80-120% limits requires a total variance of less than $148\%^2$, and meeting the MB's 85-115% limits requires a total variance of less than $44\%^2$. Stated differently, the *MB limits* corresponding to individual content uniformity limits of ±25% for a 10-actuation method *should be ±20%*, given the specified split of measurement error and between-unit and within-unit variation.

In response to Dr. Lostritto's inquiry, the Working Group mentioned that in order to understand actual variabilities fom different sources, specially designed testing should be conducted, for example, as per the protocol developed by this group earlier for the planned propsective CI experiment. Dr. Tougas stressed, however, that regardless of the actual numbers assumed in this simulation, the message about *relative* tightness of the different test protocols remains the same. This rank-ordering depends on the basic statistical priniples and not so much on the specific numbers chosen in the simulaiton. Dr. Lostitto mentioned that he had confirmed with a statistical expert at the Agency that the basic statistical analysis presented by Ms. Strickland is correct. Participants discussed that such analysis would be important if a "single-canister" DCU test needed to be developed.

Dr. Rogers asked why historic MB measurements seem to be tighter than DCU measurements. Some participants explained that companies develop CI methods so that MB falls within a tight range, for example by using multiple actuations. In addition, a broader industry dataset (IPAC-RS report to FDA) does not show a tighter range for MB. It would be important to look at the same set of data when drawing conclusions.

Dr. Lostritto asked whether extra caution used in developing the CI method would suggest that CI measurements are less variable than DCU. Other participants explained that this is not the case, because the CI test by nature involves more steps, and variabilities from each of the operational steps add up.

Ms. Strickland also explained that for mathematical reasons, the expected range of observed results depends on whether 1 canister x 10 actuations is tested or 10 canisters x 1 actuation each, all other factors being equal.

Specific Consensus Points

The Working Group agreed that:

- The acceptance criteria or action limits (terminology to be determined) for mass balance should be allowed to center off 100 % target emitted dose, to account for net systematic bias, which includes wall losses, unavoidable in any impactor. Up to 5% shift off the 100% should be allowed, as justified by applicant's scientific explanation and data.
- The range of the MB limits should be justified by the applicant, based on science and supported by data. As a default, ±15% seems reasonable.
- The term "target emitted dose" should be used instead of "label claim" when discussing limits for mass balance, because for pre-metered devices, the label claim refers to the content of the blister/capsule rather than to the emitted dose.
- It would be better to have reasonable limits and a simple test than incrementally broadening limits organized in multiple tiers.
- If a mass balance is outside expected limits, the entire set of results from that CI run should be regarded as invalid. This means that the "failing" stage requirements are also invalidated by the failing mass balance.
- In the case of a failed mass balance and consequent retesting, an emitted dose test on the same canister would not be offered as an option in the PQRI recommendations. (This would avoid the need to develop and validate a separate emitted dose test on a single canister, and would eliminate the question of single-dose vs. multi-dose product retesting).
- If upon retesting with a new canister mass balance fails with no assignable cause, an OOS investigation should be conducted.

Safe Harbor

Dr. Lostritto noted that the issue of safe harbor remains open; he expressed hope, however, that the Working Group would be able to develop its own blinding mechanisms to enable data sharing.

IV. AGREED

- Dr. Lostritto will transcribe and circulate consensus points agreed at the meeting.
- Drs. Dunbar and Mitchell will transcribe and circulate the revised, annotated flow chart for handling mass balance measurement results.
- The Working Group members will comment on the circulated drafts by email and will discuss them at the next teleconference.
- The issues of further plans and leadership of the Working Group will be discussed on the next teleconference.
- Working Group members interested in the position of Chair should contact Dr. Tougas. Dr. Tougas will clarify the process for appointing a new Working Group Chair with the Steering Committee Chair.

V. NEXT TELECONFERENCE / MEETING

The next teleconference will be scheduled via email for the week of 16 August 2004.

Finalized on 23 August 2004

EXHIBIT A

This document illustrates how emitted dose acceptance criteria compare with the mass balance criteria. The measurement error would be determined for a product via the proposed experimental design.

Total Variation=Measurement + Product

$$\sigma^2_{Total} = \sigma^2_{Measurement} + \sigma^2_{Product}$$

Product Variation=Between Unit + Within Unit (between-unit and within-unit variabilities are assumed independent of each other)

$$\sigma^2_{Product} = \sigma^2_{BetweenUnit} + \sigma^2_{WithinUnit}$$

Therefore, Total Variation = Measurement +Between Unit + Within Unit

$$\sigma^2_{Total} = \sigma^2_{Measurement} + \sigma^2_{BetweenUnit} + \sigma^2_{WithinUnit}$$

Based on Normal Distribution Theory:

99.7% of all observations lie within 3 standard deviation units of the mean.

$$L_{99.7}, U_{99.7} = \mu \pm 3\sqrt{\sigma^2_{Total}}$$

90% of all observations lie within 1.645 standard deviation units of the mean.

$$L_{90}, U_{90} = \mu \pm 1.645\sqrt{\sigma^2_{Total}}$$

Expressing the total variation in terms of Measurement, Between Unit and Within Unit Variation

$$\sigma^2_{total} = \frac{\sigma^2_{BetweenUnit}}{n_{unit}} + \frac{\sigma^2_{WithinUnit}}{n_{act}n_{unit}} + \frac{\sigma^2_{MeasurementError}}{n_{unit}}$$

implies:

$$L_{99.7}, U_{99.7} = \mu \pm 3\sqrt{\frac{\sigma^2_{BetweenUnit}}{n_{unit}} + \frac{\sigma^2_{WithinUnit}}{n_{act}n_{unit}} + \frac{\sigma^2_{MeasurementError}}{n_{unit}}}$$

$$L_{90}, U_{90} = \mu \pm 1.645\sqrt{\frac{\sigma^2_{BetweenUnit}}{n_{unit}} + \frac{\sigma^2_{WithinUnit}}{n_{act}n_{unit}} + \frac{\sigma^2_{MeasurementError}}{n_{unit}}}$$

Dose Content Uniformity Specification (The criteria used below are taken from the FDA Draft Guidance for MDIs and DPIs. The inclusion criteria refer to a sample of n = 10. This example is used here to demonstrate the relationship between MB and DCU requirements, for illustration purposes only)

$$\bar{X}_{10units,1act} \in (85\%,115\%)$$

(9 out of 10) 90% of $\mathrm{X}_{1unit,1act} \in (80\%,120\%)$

$$100\% \text{ of } X_{1unit,1act} \in (75\%,125\%)$$

Express Dose Content Uniformity Specification in terms of maximum total variation

$$\bar{X}_{10units,1act} \in (85\%,115\%)$$

implies

$$3\sigma_{total} = 115-100 = 100-85 = 15 \Rightarrow \sigma_{total} = 15/3 = 5$$

$$\Downarrow$$

$$\sigma^2_{total} = 5^2 = 25$$

$$90\% \text{ of } \mathrm{X}_{1unit,1act} \in (80\%,120\%)$$

$$1.645\sigma_{total} = 120-100 = 100-80 = 20 \Rightarrow \sigma_{total} = 20/1.645 = 12.15$$

$$\Downarrow$$

$$\sigma^2_{total} = 12.15^2 = 148$$

$$100\% \text{ of } X_{1unit,1act} \in (75\%,125\%)$$

$$3\sigma_{total} = 125-100 = 100-75 = 25 \Rightarrow \sigma_{total} = 25/3 = 8.33$$

$$\Downarrow$$

$$\sigma^2_{total} = 8.33^2 = 69.4$$

Express Dose Content Uniformity Specification for 99.7% of all individuals within 25% of LC in terms of total variation as a function of measurement, between and within unit variation. 1 unit and 1actuation per unit per result. This is the most stringent of the DCU specification.

$$\sigma^2_{total} = \frac{\sigma^2_{BetweenUnit}}{1} + \frac{\sigma^2_{WithinUnit}}{1*1} + \frac{\sigma^2_{MeasurementError}}{1} = 69.4$$

Let measurement variation be 10% of total $\sigma^2_{ME} = 6.9$

within unit variation be 40% of total $\sigma^2_{Within} = 27.8$

between unit variation be 50% of total $\sigma^2_{Between} = 34.7$

Using these values for the variation components then assuming for now that the measurement error for Mass Balance is the same as Dose Uniformity measurement error and the result comes from 1 unit and 10 actuations the total variance

$$\sigma^2_{total} = \frac{\sigma^2_{BetweenUnit}}{1} + \frac{\sigma^2_{WithinUnit}}{1*10} + \frac{\sigma^2_{MeasurementError}}{1} = \frac{37.4}{1} + \frac{27.8}{1*10} + \frac{6.9}{1} = 44.4$$

$$L_{99.7}, U_{99.7} = 100 \pm 3\sqrt{\sigma^2_{Total}}$$

$$L_{99.7}, U_{99.7} = 100 \pm 3\sqrt{44.4}$$

$$L_{99.7}, U_{99.7} = 100 \pm 3*6.66 = 100 \pm 19.9$$

Therefore the MB limits corresponding to individual content uniformity limits of ±25% for 10 actuation method would be ±20%, given the specified split of measurement error, between unit and within unit variation.

For a 5- actuation method:

$$\sigma^2_{total} = \frac{\sigma^2_{BetweenUnit}}{1} + \frac{\sigma^2_{WithinUnit}}{1*5} + \frac{\sigma^2_{MeasurementError}}{1} = \frac{37.4}{1} + \frac{27.8}{1*5} + \frac{6.9}{1} = 47.2$$

$$L_{99.7}, U_{99.7} = 100 \pm 3*6.9 = 100 \pm 20.6$$

Therefore the MB limits corresponding to individual content uniformity limits of ±25% for a 5-actuation method would be ±20.6%, given the specified split of measurement error, between unit and within unit variation.