Critique and Review of:

Report on the evaluation of the Low Dose ENFit™ Enteral Syringe Design Dosing Accuracy

CEN/TC 205/WG 16  « Catheters »

Date: 2018-06-27
1. PURPOSE

The purpose of this review is to provide perspective and statistical critique of CEN/TC 205/WG 16 - Report on the evaluation of the Low Dose ENfit™ Enteral Syringe Design Dosing Accuracy as requested by Arleni Garcia, Global Marketing Manager, BD Medical.

The stated goals of the study as shown on p.9 of the report are:

The purpose of this report is to summarize the results of the evaluation of the Low Dose ENfit™ Enteral Syringe design with regard to the following:

1.1 Low Dose ENfit™ Syringe Design Dosing Accuracy
1.2 Low Dose ENfit™ Syringe Design Applicability by Syringe Size
1.3 Low Dose ENfit™ Syringe Design Capability versus Currently Marketed Enteral/Oral Syringe Designs
1.4 Low Dose ENfit™ Syringe Design Capability versus Regular ENfit™ Enteral Syringe Design
1.5 Substantial Equivalence of Low Dose ENfit™ Syringe Design to Currently Marketed Enteral/Oral Syringes

This report will also analyze the dosing accuracy target requirements as stated by industry versus the actual dosing accuracy of currently marketed syringes where dosing accuracy refers to the dosing accuracy of the syringe when used in combination with a compatible nasogastric feeding tube to administer medication/nutritional media to a patient.

The scope of the evaluation included test and administration devices as supplied by the individual device manufacturers.

2. SAMPLE LIMITATIONS

The limitations caused by the number and sampling of the parts used in this study are discussed first. These limitations are independent of the data analysis and its interpretation.

The most common sample size used in this study was 16 syringes for each condition studied. The conditions involved a number of syringe or testing characteristics including:

1. Syringe manufacturer
2. Syringe size
3. Syringe design such as ENfit™, Nutrisave2 or Reverse Luer Tip
4. Disconnected with feeding tube connector oriented up or down
5. Fluid was administered through the feeding tube with the feeding tube connector up or down
6. The syringe was filled from a cup or a straw

In the comparison of testing and administering characteristics, the sample size of 16 may have been appropriate because the goal was comparison of the accuracy of the different measurement or delivery techniques. These procedures were practiced consistently within the study and comparisons are accurate. However, to evaluate and compare manufacturers, sizes and designs, a sample size of 16 syringes provided by the individual device manufacturer has serious limitations for three main reasons.
A. Sampling Variation in a 16 Part Sample

To illustrate the sampling variation with a 16 part sample, consider the results from the Halyard Health Test Report for the 5ml ENFit Syringe as shown on p16 of the report and reproduced below. The graphical view of the data with a +/- 10% Accuracy Requirement limits as shown on p 18 of the report is also reproduced below.

The results show the number of failures with the Connector DOWN is 0 (0%) and this meets requirements. The results also show the number of failures with the Connector Up is 5 (31%) and this does not meet the requirement. However, if we assume the parameter estimates for the mean and standard deviation for the populations were accurately estimated as shown above, we can simulate the results from taking 16 samples and visualize the variation in our Number of Failures result. Results for 1000 simulations are shown below.
Here we see that the results for 16 samples with the Connector DOWN could have been anywhere from 3 to 10 failures where the actual test results were 5 failures, fairly close to the expected failure rate of 5 or 6. On the other hand, simulation results for Connector Up could have been anywhere from 0 to 6 failures where the actual test results were 0 failures. This result is far from the expected value of 3 failures which would have resulted in a “No” to the expectation of meeting requirements whereas the 0 value resulted in a “Yes” to the question of meeting requirements.

Every sample size has an associated variability usually expressed as a confidence interval, but shown here visually for simplicity. If the sample size used for this study was 32 instead of 16, the expected range in failures would have decreased, and again for 50 or 100 progressively. Every study must draw the line at some point. The question for the researchers here is; “Does the range of variation caused by selecting 16 for a sample size meet their goals and expectations for the study?”. This is a question only they can answer.

**B. The Sample Taken does not Represent the Population Conclusions are Drawn On.**

This limitation is not a result of the sample size of 16, but more of the sample itself. It is a fatal flaw for the conclusions of this study.

Consider manufacturing results shown below for a typical injection molding process where a part diameter for a catheter, syringe, tubing, etc. is measured over time. In nearly every manufacturing process, the
average output drifts, shifts and trends resulting in measurements similar to those seen here. This data was collected over just 10 hours from one manufacturing line. Clearly, the four red boxes indicating four different 16 sample snapshots of the process would yield completely different views of the distance of the diameter to target from Design 1, Manufacturer A, etc. These changes in output are seen in all manufacturing processes and allowed as long as the product meets specifications.

This typical variability in the average output over 10 hours is magnified when one considers days, weeks and months of time, injection mold changes, shift changes, multiple manufacturing lines, seasonal changes etc. In our study, the mean accuracy (%) will change depending on when the parts were taken and this will strongly impact results. The authors compensate for this by noting in the Scope of the study that the syringes were provided by the individual manufacturers, essentially making them responsible for the results from their parts. However, this does not assure the parts provided represent the population of that manufacturer, design or volume.

**C. Results are Dependent on the Average of the Sample which is Highly Variable**

The graph below shows the averages of the data from the injection molding process divided into 16 part samples and the average for the subset is taken. The graphic indicates the average of the sample can vary from .319 to .333 ( +/- 2.3%). This variation is independent of the sampling variation described in part A, of this section and has it’s own detrimental effect on our trust in the results. The calculations in all three main analyses in the report are dependent on the mean value of the 16 parts measured. For example; the Number of Failures calculation changes from 2 to 6 as the mean value increases from 5.4 to 8.8% for the 5ml syringe samples. The 5.4 to 8.8% was the range of mean dose accuracy (%) seen for the 5ml syringes described earlier. This dependence of the analysis results on the mean of the sample, which is highly variable, casts a doubt on the results throughout the study.

![Graph showing variability in part diameter measurements](image)
3. Review of: **DETERMINATION OF LOW DOSE ENTERAL SYRINGE DESIGN**
APPLICABILITY BY SIZE TESTING SUMMARY

The purpose of this study as described on p 16 of the report is to progressively test smaller and smaller syringe volumes to determine the point at which the syringe volume size would benefit from the low dose syringe design compared to the standard design.

The measurement used to make this determination was the Number of Failures > 0 with a failure being a dose measurement outside + / - 10% of the target value. The summary of this work is shown in the table below from p 16 of the report.

<table>
<thead>
<tr>
<th>Syringe Size</th>
<th>Fill Volume (mL)</th>
<th>Accuracy Requirement (+/-) mL</th>
<th>Administration / Disconnection Orientation</th>
<th>Dosing Accuracy (Min)</th>
<th>Dosing Accuracy (Max)</th>
<th>Dosing Accuracy (Avg)</th>
<th>Dosing Accuracy (SD)</th>
<th># of Individual Failures</th>
<th>Meets Requirement (Yes/No)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 mL</td>
<td>2.4</td>
<td>0.24</td>
<td>FT Connector Down</td>
<td>-0.055</td>
<td>0.120</td>
<td>0.026</td>
<td>0.046</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>12 mL</td>
<td>2.4</td>
<td>0.24</td>
<td>FT Connector Up</td>
<td>-0.017</td>
<td>0.117</td>
<td>0.044</td>
<td>0.042</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>10 mL</td>
<td>2.0</td>
<td>0.2</td>
<td>FT Connector Down</td>
<td>0.016</td>
<td>0.178</td>
<td>0.098</td>
<td>0.042</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>10 mL</td>
<td>2.0</td>
<td>0.2</td>
<td>FT Connector Up</td>
<td>0.063</td>
<td>0.175</td>
<td>0.115</td>
<td>0.032</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>6 mL</td>
<td>1.2</td>
<td>0.12</td>
<td>FT Connector Down</td>
<td>-0.071</td>
<td>0.038</td>
<td>-0.018</td>
<td>0.035</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>6 mL</td>
<td>1.2</td>
<td>0.12</td>
<td>FT Connector Up</td>
<td>-0.099</td>
<td>0.101</td>
<td>-0.013</td>
<td>0.056</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>5 mL</td>
<td>1.0</td>
<td>0.1</td>
<td>FT Connector Down</td>
<td>-0.048</td>
<td>0.099</td>
<td>0.054</td>
<td>0.044</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>5 mL</td>
<td>1.0</td>
<td>0.1</td>
<td>FT Connector Up</td>
<td>0.011</td>
<td>0.153</td>
<td>0.088</td>
<td>0.037</td>
<td>5</td>
<td>No</td>
</tr>
<tr>
<td>5 mL</td>
<td>0.6</td>
<td>0.06</td>
<td>FT Connector Down</td>
<td>0.075</td>
<td>0.075</td>
<td>0.075</td>
<td>n/a</td>
<td>*1</td>
<td>No</td>
</tr>
<tr>
<td>5 mL</td>
<td>0.6</td>
<td>0.06</td>
<td>FT Connector Up</td>
<td>-0.005</td>
<td>0.089</td>
<td>0.032</td>
<td>0.041</td>
<td>*1</td>
<td>No</td>
</tr>
<tr>
<td>1 mL</td>
<td>0.2</td>
<td>0.02</td>
<td>FT Connector Up</td>
<td>-0.004</td>
<td>0.060</td>
<td>0.025</td>
<td>0.020</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>1 mL</td>
<td>0.2</td>
<td>0.02</td>
<td>FT Connector Up</td>
<td>-0.002</td>
<td>0.081</td>
<td>0.030</td>
<td>0.024</td>
<td>10</td>
<td>No</td>
</tr>
</tbody>
</table>

*testing stopped after a single failure was noted

There are three reasons the data in this study is suspect making the results inconclusive.

1. The average dose for every test (shown in the graph below) except the 6 ml volume, both tip up and down, were positive indicating a bias in the measurements taken. The alternative cause is that manufacturers purposely produce parts that deliver on the high side for some commercial reason. Given the impact of the mean accuracy % of the sample on the Number of Failure results as described earlier, this bias has a strong impact on the Number of Failures.

2. The change in the mean accuracy (%) by syringe volume also shown in the graph below, again indicates that the mean accuracy % varies by the sample provided which then is key to the Number of Failures results. For example, the 5ml syringe data where Connector Down results in 5 failures and does not Meet Requirements whereas Connector Up results in 0 failures and does Meet Requirements. Since the Up / Down configuration does not affect the mean accuracy, the difference must be in the means of the 16 part samples tested.

3. Data were not tested for outliers. For example, the 1 ml fill wts as taken from p 171 of the report are shown graphically below. The boxplot indicates two outliers in the data for the Connector Up. The time series plot of the time order of the data also suggests a physical reason for the outliers. After each outlier which is a spike upward, the next data point is a strong move downward suggesting a human adjustment to the special cause variation that caused the outlier. This is common with data resulting from human procedure steps and then measurement. The study report simply calls these “failures”, the same as the other three points beyond the 10% accuracy limit in this group of 16 parts.
Page | 8 - Review of “Evaluation of Low Dose ENFit™ Enteral Syringe Design Dosing Accuracy”
There are two issues with the data analysis and interpretation for this study. These issues are independent of the problems with the data itself described earlier.

1. The conclusion on p 19, "The feeding tube connector “DOWN” orientation typically had slightly lower delivered volumes than the UP” orientation. " is just a descriptive observation. To statistically test UP vs DOWN, a two-way analysis of variance should be carried out with the first variable; Syringe Volume and the second variable; Connector UP/DOWN. The most likely result will be that Connector will be insignificant. This is helpful because it will allow combing the 32 samples for each syringe volume (both UP and DOWN) giving a more powerful comparison.

2. The response in this study is highly variable and has extremely low power. Having one failure in 16 samples render the syringe volume incapable of meeting accuracy requirements and zero failures in 16 indicating the syringe volume is capable is a poor choice. Choosing to convert continuous measurements to proportions is always poor practice. When the proportion is based on 16 samples, it is extremely poor. This was seen in the results when the 5 ml sample failed for the UP configuration and passed for the DOWN configuration when the configuration overall had no effect on the mean dose accuracy (%). It was also seen when the 5 ml syringe volume failed but the 6 ml syringe volume passed. If the test was repeated on a new set of samples from each volume, one would expect a much different result.

My observations from this study are that:

1. Results on the number of failures per 16 samples (or mean dose accuracy (%)) in each volume are meaningless because the samples used vary in their mean accuracy (%) based on manufacturing process variation and they do not represent the sample population we are drawing conclusions about.

2. Later in the report, the findings from this study will be used to support that concept that the 1 ml syringe is the “worse case scenario” because it had the highest number of failures. I believe that it is actually the best case scenario because its dose accuracy standard deviation was about .02 whereas all the other syringe volumes had a standard deviation of about .04. See table above. The only reason there are increasing failures with decreasing syringe size is because the + / - 10 % accuracy limits grow increasing smaller with the smaller target volume.

3. The statistical analysis comparing Connector Tip UP / DOWN is the only valid information available in this study if the proper two-way ANOVA analysis is used because the comparison will be on a syringe volume basis and will incorporate all the data in the study. However, even if a significant difference is seen, it will be so small as to be physically unimportant.
4. Review of: BASLINE EVALUATION OF CURRENTLY MARKETED ENTERAL SYRINGE DESIGNS

The purpose of this study as shown on p19 of the report is “to baseline currently marketed enteral syringe designs to determine dosing accuracy performance as it relates to the following requirements:

Dosing accuracy target of +/- 10% of a 0.2mL dose delivered from a 1mL syringe, where dosing accuracy is defined as the actual dose delivered to the patient side (into the feeding tube).”

The data collection was setup as a full factorial with 5 factors and 16 parts in each test condition. The five factors were:

1. Syringe Manufacturer and Design – 6 levels – Five manufacturers and Reverse Fit Design
2. Disconnect Syringe Tip Orientation – 2 levels – UP / DOWN
3. Administer Syringe Tip Orientation – 2 levels – UP / DOWN
4. Fill Technique – 2 levels – Straw / Cup
5. Low Dose Design – 2 levels – Low Dose Design / Std Design

Limitations of the Parts Selected and Sample Size of 16 Parts

1. The data collected in this study suffer from the same issue described for the last study in that parts tested do not represent the population we are trying to draw conclusions about. The mean dose accuracy (%) of the subgroups of 16 will vary due to the variability in the manufacturing process that produced them. As such, comparisons between the Syringe Manufacturer and Design (Variable 1) and comparison of the Low Dose Design (Variable 5) are not valid. Unfortunately, this was the stated purpose of the study.

2. The limitations of the sample size of 16 are small. Given the variation in the data within a condition, a sample size of 16 will have sufficient power to detect physically small differences in conditions which will be valuable to the researcher. Based on the power curve shown below which assumes a full factorial with 5 factors and a typical standard deviation of 5% (from p 26 of the report), an effect size of 2% accuracy or greater will be detected with 84% probability when using 16 parts per condition. This is generally accepted as a reasonable experimental power.
There are four issues with the data analysis and interpretation for this study. These issues are independent of the problems with the data itself described earlier.

1. The ANOVA analysis of the data is inappropriate because it does not discriminate between differences due to manufacturers, low dose design or reverse orientation tip design which are of key interest to the researcher and the tip orientation variables and cup vs straw filling effects which are of secondary interest. Not only is there no prioritization, but combining the levels of 5 variables as ANOVA category levels completely confuses any possible conclusions of substance. The Connecting Letters Report on p 30 of the report which compares mean accuracy (%) values between test conditions is shown below.

![Connecting Letters Report](image)

The conclusion from this output is “There are significant differences in the means and variances of the tested syringes” which is not only not very insightful, it is incorrect. This may explain why it is called an “Observation”.

If representative parts had been collected, the proper analysis of this very exhaustive study would have been to treat it as a full factorial in five factors as described by Box, Hunter and Hunter (2005). This analysis would have allowed for the determination of all main effects of the five variables, but also interaction effects. For example, is the change in mean dose accuracy (%) caused by cup fill vs straw fill for the low dose ENFit™ design also seen in the standard ENFit™ design.

The current data can be analyzed appropriately for the effects of variables 2, 3, and 4 (the measurement procedure variables) and their interactions could be estimated. In this analysis, comparisons would be made within each syringe manufacturer / design and pooled to yield one result for all the data. Apples – to – apples. This would allow the researcher to at least learn the measurement factors that have an effect on the results so that future studies could be designed more efficiently and appropriately.

2. The ANOVA analysis assumes that variances are equal between groups. Here they clearly are not as shown in the summary table on page 26 of the report. So the ANOVA analysis does not meet that assumption.
Interestingly, one analysis of the current data that would yield useful results would be to study the variability of the response as opposed to the mean dose accuracy (%). Unlike the mean, the standard deviation of the output from production processes are reasonably stable. Therefore, analysis of the standard deviation at each condition to determine which manufacturers, designs and measurement procedures yield the highest and lowest variability would provide insight as to how to achieve the best process capability by understanding one key driver; variability.

3. There is no analysis of the data for outliers and their impact on the results. Analysis of the data as a full factorial would allow for analysis of residuals to identify outlier data points and remove them if appropriate.

4. Comparing the Low Dose ENFit™ design to other manufacturers and designs in the “worse case conditions” has no rationale. As described on page 27 of the study report, comparison of the Low Dose design to others was done only for worse case conditions. The rationale for this is not clear. What conditions represent “worse case” conditions were also not explained.

The negative impact of this approach is that nonequivalence is most difficult to reject when the data have the most variability. This is same as rejecting the null of a two-sample t-test. It is most difficult to reject when the data variability is highest. Comparisons under worse case conditions is essentially “stacking the deck” for equivalence.

This anticipates the impact on the equivalence testing that should have been performed. In fact, no equivalence tests were carried out. This means that the conclusions on p30 of the report, shown below, were not support by the correct analysis of the data. For this reason, they are called “Observations” rather than conclusions support by statistical analysis of the data.

**Significant Observations**

- There are significant differences in the means and variances of the tested syringes.
- Standard EO syringes (male tip) typically showed better dosing accuracy than reverse orientation EO syringes (female tip).
- Standard ENFit™ design syringes (syringe A) perform similarly to one of the currently marketed reverse orientation EO syringes (syringe F).
- Low Dose ENFit™ design syringes perform similar to currently marketed standard EO syringes.
4. **Review of: Capability Analysis of 1 mL Syringe Dosing Accuracy % (Using +/- 10% target limits)**

The purpose of this analysis is to compare process capability of different syringe designs and manufacturers. The analysis is described on p30 of the report.

“Capability analysis was performed for each syringe design and fill method to determine the capability of each syringe design/method to meet the dosing accuracy requirement of +/-10%. Capability was determined using the pooled data for each syringe fill method and overall performance. Reverse orientation syringes were analyzed individually and as a group due to unique differences in tip design; standard orientation syringes do not have significant differences so were reviewed only as a group.”

Results of this analysis are inaccurate and cannot be interpreted for the following two reasons.

1. The mean dosing accuracy (%) for each population is not reflected in the 16 part samples. Since the mean is a key component of the process capability calculations, these estimates are not accurate.

2. Samples from different populations were incorrectly combined. For example, the statement in the paragraph above that standard syringes do not have significant different differences so were analyzed as a group. This was never shown in the data analysis and is incorrect. Combining data from different populations will increase the standard deviation and therefore reduce the capability estimate.

In my estimation, this section and analysis are woefully incorrect and not useful.

---

**References**