November 18, 2015

Division of Dockets Management (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Rm. 1061
Rockville, MD 20852

Re: FDA-2015-D-2537
FDA DRAFT Guidance for Industry: Request for Quality Metrics

Dear Sir/Madam,

Thank you for this opportunity to provide comments on the FDA "Request for Quality Metrics" Draft Guidance for Industry. My comments are intended to objectively assess the potential outcome of the draft guidance implementation versus the intended outcome, as well as potential impact to the Agency, industry and patients.

I firmly support the use of risk-based decisions across the Agency and industry to identify how best to use available resources to protect and promote patient safety in a way that is commensurate with the need. Additionally, I fully support the intended multi-faceted purpose of this draft guidance to: (1) guide risk-based resource allocation decisions within the Agency, (2) improve the Agency's ability to predict drug shortage circumstances, and (3) encourage the implementation of state-of-the-art innovative quality management systems across the pharmaceutical industry. Therefore, my comments do not question the merit of that intended purpose. However:

I do not believe that the proposed call-for-data approach will achieve the intended multi-faceted purpose of the draft guidance, and even more concerning will divert Agency and industry resources from activities critical to promoting and protecting patient health.

The basis for my disagreement with implementing the proposed call-for-data approach as a mechanism to achieve the purpose of the draft guidance is as follows:

1. Assessment of data out-of-context will result in unsubstantiated conclusions and unintended consequences. Assessment of data without the context of the surrounding circumstances renders the conclusions unreliable until verified with supporting evidence. Therefore, basing decisions on unsubstantiated conclusions has inherent risks that impact patient safety. Additionally, consequences will occur that are unintended by the Agency, such as inconsistency of interpretation, driving the wrong behaviors, and gaming.

   There are multiple examples to illustrate how out-of-context data could be detrimental:

   a. Varying complexity of product formulation and process sophistication across the industry could result in a lower number for any given metric for any given company. Imagine that the low number is actually world class for the operation in question. The low number out of context,

however, would lead one to assume that the cause of the low number was related to poor performance, and the company would be viewed as higher risk.

b. Taking the example in 1.a. above and modifying it to compare the exact same product across multiple companies would still lead to unsubstantiated conclusions. For example, low data could be the result of a lower risk tolerance of one company compared to another, even though they both may be experiencing the same risks to their products. Without context, one would assume that the company with the seemingly better data (higher risk tolerance) would be of lower risk. The draft guidance would unintentionally reward this behavior.
c. Two companies with similar performance data may investigate and mitigate risks with varying levels of rigor, but the commitment to prevention might not be apparent upon reviewing the data out of context. Comparative conclusions would therefore be uninformed in the absence of context, and the companies would be viewed as comparable in risk.
d. Understanding root causes for why data is at a certain level is critical to drawing conclusions. Shifts in metrics could result from many factors that are outside of the product and/or process itself. For example, equipment could have malfunctioned, a facility excursion might have occurred, unexpected downtime due to employee staffing might have occurred, delays in moving material through the facility might have caused material to go beyond its hold time, a spill might have occurred, etc. Again, without context, one would conclude that a lower number signals higher risk or poor product quality.
e. The overall process of calling for isolated data out-of-context could signal to less mature companies that these specific metrics are the most important to track. However, every company and every operation is different. It is imperative for each company to evolve its assessment of metrics as the performance of its products and processes evolve.
f. FDA inaction as a result of data review might signal to less mature companies that FDA approves of the company's overall performance, and the company, therefore, might not strive to continuously improve.
g. In order to determine risk to patient safety through risk to product quality, mature organizations assess data through a complex aggregation of (1) assumptions, (2) known history, (3) recognition of known-unknowns, (4) risks originating from robustness of the process, tolerance of product variability, and knowledge of the supply chain, (5) trending of failure root causes, (6) in combination with other metrics to provide a holistic assessment, etc. Mature organizations aggregate the data in an effort to identify correlations that could lead to product quality predictability. Calling for single metrics to review in isolation would be beneath the sophistication of these companies, and could distract these companies from dedicating its resources to more sophisticated analyses. Additionally, it may unintentionally suggest to less mature companies that isolated metric review is what the Agency expects of companies.
h. Companies have varying levels of risk tolerance based on the culture of that company, past performance, status with Regulatory Authorities, etc. Data taken out of context might drive the unintended consequence of companies accepting greater risk so as to avoid a poor metric. The draft guidance would unintentionally reward this behavior.
i. Some companies may focus on managing the metrics (i.e. the numbers) rather than resolving the issues. The draft guidance would unintentionally reward this behavior.
j. Some companies (albeit a small percentage) may intentionally game the metrics in order to give the appearance of improvement, and therefore be viewed as lower risk by FDA. The draft guidance would unintentionally reward this behavior.

Below is an assessment of how each proposed metric could lead to unsubstantiated conclusions and unintentional consequences without understanding the surrounding circumstances:

- **Lot Acceptance Rate**
  - Company A has a better lot acceptance rate than Company B because Company B has a lower risk tolerance and rejects batches based on potential risk to product quality. Company A only rejects batches that actually fail specifications, and even under those circumstances inappropriately brackets portions of the batch for release. Company A would appear to be of lower risk based on the data alone.
  - Companies A and B interpret the metric differently for which rejected batches to include. For example, Company A has filed in-process specifications, so batches failing these specifications are rejected and are included in the calculation. Company B only uses finished specifications as the basis for rejection. Company C additionally includes batches that have been released to the market, but now are found to fail on stability, whereas Companies A and B do not.
  - The data will be difficult to compare based on volume differences. One rejected batch out of one attempted is considerably different than one rejected batch out of 100 attempted. Also, the volume will vary from month to month, and year to year, which would incorrectly give the appearance of lot acceptance rate shifts. Additionally, this does not take into account the complexity of the formulation and process.

- **Product Quality Complaint Rate**
  - Company A launched a product for the first time that serves the elderly population. Known psychology studies have shown that loneliness drives higher complaint rates from the elderly population. Based on the data alone, Company A would appear to be of higher risk than a company that does not serve this same population.
  - Complaint data can be received from a lot that has been on the market for 1.5 years, but yet the denominator is based on products released to the market in the current time period. This is a common calculation for complaints, but companies are able to contextualize shifts based on known shifts in release volume. Shifts without the context would be viewed as impacting risk.
  - Some products are known to receive higher complaints due to therapeutic category, such as pain medications. Companies releasing these types of products would appear to be of higher risk than other companies.
  - Interpretation of when to "count" a complaint will vary. Companies employ varying triage methodologies for when complaints are received versus when they are counted. There are certain portions of complaints that are not valid. Based on when companies begin "counting" their complaints (before, during or after triage) will render the numbers meaningless without context.

- **Invalidated Out-of-Specification (OOS) Rate**
  - Company A does not declare an OOS result until experiments prove the result is valid. Therefore if invalid, then the OOS designation was never given to the failure, and therefore, there is no record of invalidating an OOS result. Company B declares the OOS

result as soon as it is obtained. Company A would be rewarded with a lower risk designation.
    - This metric is not important in and of itself. It is a vital pathway to follow on inspection, because it gives investigators insight into the level of quality control in the company. For example, on inspection investigators often ask for a list of all OOS investigations that were invalidated. Upon reading the investigation report, the investigator can determine if a company is rigorously investigating the failure, or if it is quick to tag the failure with an unsubstantiated laboratory error in order to inappropriately allow for release of the product.
    - More important than this metric, the mark of a good laboratory includes one that has rigorous method validation and transfer results, low human error rates, and scientifically justified root cause determinations.

- **Annual Product Review (APR) or Product Quality Review (PQR) on Time Rate**
    - This metric could possibly give the Agency some simple information and might be useful for the following reasons:
        - If the Agency modifies this metric to only assess the on-time rate, then no additional reporting would be needed from industry (it is not clear to me why comparing the number of APRs completed on-time versus the number of products produced by the establishment would be important).
        - I like this metric for the simple reason that it is a CFR requirement, but yet companies are missing the date, or not submitting the report at all. This alone gives the Agency an indication of how serious the company is about regulatory requirements, and can be used by the Agency to justify allocating resources to inspect companies that are not submitting their APRs. Simple.

2. None of the proposed metrics will provide a predictive mechanism for drug shortages. The identification of risk for drug shortages requires a sophisticated assessment of aggregated information. For example, in order to identify the potential for drug shortages, the following information is critical to understand:
    a. Inventory levels at the finished product manufacturer's facility and the inventory of suppliers throughout the supply chain.
    b. Supply chain sophistication (robustness of the capacity of the suppliers chosen, management of duplicity in the supply chain (dual sourcing does not automatically equate to lower risk), supplier delivery performance, supplier quality performance, communication of expectations by the manufacturer to the suppliers, etc.).
    c. Geographical risk of supplier location as it relates to the economic stability of that region of the world, the existence of a strong regulatory body in that region, as well as a higher potential for natural disasters.
    d. The ability of the manufacturer to generate a stable and accurate forecast.
    e. The commitment of the company to recognize that when sales and profits are not sufficient to reinvest in the quality systems, facilities and operations that the company then makes the business decision of whether or not to continue manufacturing the drug in question. If the decision is to not continue, then this could lead to an out-of-stock situation on the market and eliminate patient access (requires an understanding of the inventory of other

manufacturers). If the decision is to continue, then profits from other products must be used to maintain operations at the expected levels.

3. The resource burden provided in the Federal Register is grossly underestimated for both the industry and Agency. The following is a breakdown of why I believe the resource burden is grossly underestimated:
    a. **Agency:** collecting data from industry into a repository is not time intensive, and might follow the estimations outlined in the Federal Register. However, the Agency would at a minimum need to identify how to compare the data across companies in order to determine where best to allocate its resources. In order to do this, the Agency would need to segregate companies by product and process complexity and risk, then determine how to rank companies into those categories. Many companies, however, have multiple levels of product and process risks, so the Agency would need to determine how to triage this risk within a company in order to conduct cross-company comparisons. This would prove time intensive if not impossible, especially in light of the constantly evolving industry (new product introductions, acquisitions, site changes, etc.), as well as inaccurate/incomplete data the Agency has on establishment information. Even if the Agency could accomplish the above, the task of putting the data into context would be insurmountable. The Agency would have to pull data on each company from many databases across branches to even partially understand the known risk of the company (inspectional history, recalls, enforcement action, product approvals, etc.) to match with the call-for-data proposed in the draft guidance. Establishing correlations would be time intensive if not impossible, especially in light of the arguments presented in #1 above that demonstrate the need to review the data on-inspection/in-context.

    b. **Industry:** data gathering and metrics reporting varies widely across the industry, and even across plant sites within a single company. Work would be needed by industry to gather the data in the prescribed way by the draft guidance. Additionally, mature companies may have the empirical data feeding into more complex calculations through automated mechanisms. This could require recoding of the software to gather the data as requested in the guidance. Some companies still operate on a paper-based system. The data gathering in all cases would require resources on an ongoing basis to pull the data in the way desired by the Agency. Some companies may not be assessing the data requested in the draft guidance at all. For example, tracking invalidated laboratory results may not be tracked at all, but rather, more importantly, tracking may be conducted for root cause trends (OOS invalidation is, of course, not a root cause). Additionally, any data reported to a regulatory agency typically goes through a rigorous review process and verification to ensure complete accuracy. This would be an ongoing resource burden. And finally, transferring the data in-house to an automated FDA system will require ongoing resources. The larger concern is that the added burden (no matter now large or small) would result from work I do not feel will enable the Agency to achieve the multi-faceted purpose of the draft guidance, and would not reward companies that are already performing within and beyond Agency expectations.

Opinions Related to the Optional Metrics Provided in the Draft Guidance

Overall, I feel the optional metrics provided in the draft guidance have some merit that could enable the Agency to justify how to allocate its resources. That merit is as follows:

- By requesting a response to *any* question, the Agency will be able to justify dedicating its resources to those companies that do not respond at all. This immediate segregation will reward companies that are participating and will highlight the risk of those companies that do not acknowledge the authority of the Agency.
- The optional metrics in the draft guidance require binary responses, and as such, the reporting burden on the industry would be much lower than the estimates provided in the Federal Register.
- By asking the industry if the three activities related to the optional metrics are being conducted, then it signals to the industry the importance FDA places on those activities. It could *help* achieve the third purpose of the draft guidance: encourage the implementation of state-of-the-art innovative quality management systems across the pharmaceutical industry.

Factors for the Agency to consider related to the optional metrics are as follows:

- The responses to any binary questions will become static once the desired state is achieved. For example, if Senior Management is engaged in the APR review process, there is no likelihood that a company would change their response from a "3" to a "1", "2" or "4" in the future.
- I believe the Agency already has data in hand that would allow it to justify focusing its resources on a specific segment of the industry. For example, companies that are not submitting Annual Reports as required by 21 CFR 314.81 should already be viewed by the Agency as higher risk. The Agency should be able to justify focusing its resources on that segment of the industry, as well as companies that "grossly" miss the reporting due dates ("grossly" to be defined by the Agency). This could be combined with data the Agency already has on the inspectional history (performance, as well as whether an inspection has been conducted at all), enforcement history, and recall history of those companies. Using the information already available to the Agency would result in zero reporting burden for the industry.
- The optional metrics still require context to draw a substantiated conclusion regarding the risk of the company. Additionally, there is always a concern that a certain (albeit small) percentage of the industry may falsify the responses.

Assessment of each Optional Metric:

1. Senior Management Engagement. Although having senior management signatures on the APR is a level above not having their review at all, it is not possible through this question to ascertain the rigor of that review and if appropriate actions were taken as a result of the review. This question, however, would signal the importance the Agency places on having senior management involved and would drive less mature companies to begin including senior management. More mature companies already engage senior management, and the reporting burden would therefore be minimal.

Marla A. Phillips, Ph.D. • Director • Xavier Health
Xavier University • www.Xavier.edu/Health • 3800 Victory Parkway • Cincinnati, OH 45207

2. CAPA Effectiveness. Although the Agency desires to push less mature companies in the industry into more rigorous root cause analyses, I do not believe this metric will drive that intended outcome. Also, by asking for the percentage of occurrence of the wrong behavior (i.e. CAPA involving retraining), this does not signal to the industry what the Agency feels the industry should be doing. Although not ideal, better metrics to help achieve the intended outcome would be to ask: (1) is your organization conducting a rolling root cause trend analysis for each investigation? If yes, then (2) is your organization also conducting a year to year comparison of root cause trends? An additional or optional question might be to ask for the training completion rate of the company. Of course, the ultimate test of CAPA Effectiveness is the reduction of repeat failures. However, I believe this can only be determined on inspection, and therefore, in context.

3. Process Capability/Performance. I believe mature companies should be using statistical analyses for many factors, including critical quality attributes, in-process testing and stability testing, as well as requesting the process capability of their suppliers of incoming materials and that of their contract manufacturers. Statistical analyses can inform decisions and trigger preventative actions. Asking the question proposed for this optional metric would be a low reporting burden for the industry and would signal to the industry the importance the Agency places on statistical analyses.

Ideas for Agency Consideration

In an effort to provide alternative ideas for Agency consideration, I am providing the following comments and the attached proposal. The following is a list of efforts I feel can make a difference:

1. The Agency can focus its resources on achieving mutual reliance with respected regulatory authorities in other regions of the world. This will allow the Agency to utilize the inspectional findings of other regulatory authorities for companies the FDA cannot currently reach due to resource constraints. The FDA could then assess those inspectional findings and render its own opinion of the compliance status of the company and any enforcements actions it determines is necessary. Importantly, members of the pharmaceutical industry expressed during the FDA/Xavier PharmaLink Conference that they:
    a. Do not necessarily want a reduced inspectional frequency by the FDA, and view it as an opportunity to ensure operations are on track with FDA expectations.
    b. Would, however, like to see fewer inspections in the same year by multiple regulatory authorities – i.e., they would like to see mutual reliance where one agency would conduct an inspection in lieu of all applicable agencies conducting inspections in the same year.
    c. Would like to see the playing field leveled such that their competitors who are currently not inspected at all, are also inspected.

2. The Agency could focus its resources on categorizing each inspectional finding by criticality. Currently inspections are ranked using NAI (No Action Indicated), VAI (Voluntary Action Indicated) and OAI (Official Action Indicated). Also, the individual findings are listed in priority order. However, neither gives an understanding of the risk ranking of each finding relative to clinical

relevance. For example, the highest priority finding at an excellent facility might be of very low *actual* risk. It is recommended that the inspectional findings be ranked by risk (i.e., Critical, Major, Minor) based on impact of the failure to patient safety. This would allow FDA to more easily compare inspectional outcomes across companies, and would enable senior management to better understand the criticality of the findings.

3. I propose that the Agency identifies high risk companies using information it already has in hand, then utilizes a set of metrics to review on inspection that holistically gives an indication of product quality risk, instead of conducting a call-for-data. Agency resources could be shifted from the low risk companies to the high risk companies, again by using information already in hand, or at most, requesting responses to a couple of binary questions as discussed above. It would be ideal to add zero burden to the mature companies that have been performing at or above Agency expectations already. Xavier University led an initiative with 30 industry professionals from August 2014 – June 2015 to identify a roadmap of metrics across the total product lifecycle that could provide a holistic picture of product quality risk. Attached to these comments is that proposal for Agency consideration and industry use.

Thank you for your time and consideration of these comments.

Sincerely,

Marla A. Phillips

Attachment: "Xavier University Pharmaceutical Quality Metrics Proposal"

Marla A. Phillips, Ph.D. • Director • Xavier Health
Xavier University • www.Xavier.edu/Health • 3800 Victory Parkway • Cincinnati, OH 45207