To: Sunny Balwani[sbalwani@theranos.com]
Cc: Elizabeth Holmes[eholmes@theranos.com]
From: Daniel Young
Sent: Sat 4/12/2014 3:33:35 PM
Importance: Normal
Subject: RE: Follow up to previous discussion
Received: Sat 4/12/2014 3:33:37 PM

I wanted to mention again that Tyler's concern about PT was raised right after we learned that CLIA had split the actual PT sample they had received. As I had explained to Tyler, this was the wrong thing to do, as I had also explained to Mark, Adam, Langly, and Hoda after we learned about this (recall we also had that lengthy email exchange about it). Also at that time, Adam and Mark were weighing in on similar legal interpretations regarding PT. What CLIA did at that time went against normal PT procedures, as well as our own internal procedures. (By the way, the results obtained when they split the PT sample were not meaningful, were discussed outside of CLIA, and caused confusion with some people, including Tyler at that time.)

I had discussed the proper PT and AAP procedures with Tyler twice when this all came up – and considered his questioning to stem from the questioning of Adam and Mark.

As I noted below, the last time I spoke to Mark about PT, he was still not on the same page regarding how we plan to do PT and AAP.

-Daniel

From: Sunny Balwani
Sent: Saturday, April 12, 2014 12:12 AM
To: Daniel Young
Cc: Elizabeth Holmes
Subject: Re: Follow up to previous discussion

Wow. I just read his email to u. You should escalate emails like this to us ASAP. He was actually questioning the legality of our PT process? Deeply disgusting. I will address this along with rest of his insults.
On Apr 11, 2014, at 11:25 PM, "Daniel Young" <dyoung@theranos.com> wrote:

> My comments are below in red. Please let me know if you have further thoughts or want to discuss any of this. Thanks.
>
> ---
>
> **From:** Elizabeth Holmes
> **Sent:** Friday, April 11, 2014 4:35 PM
> **To:** Daniel Young
> **Cc:** Sunny Balwani
> **Subject:** FW: Follow up to previous discussion
>
> Take a look at this. Let me know where all this data is from/what the data is, whether you had exchanges with him in which he forwarded marketing articles, and also comments line by line on the below.
>
> Separately re: marketing articles, I believe you already talked to him about the fact that what people are writing about our infrastructure is its longitudinal power in the context of lab to lab variability?
>
> ---
>
> **From:** Tyler Shultz
> **Sent:** Friday, April 11, 2014 3:38 PM
> **To:** Elizabeth Holmes
> **Subject:** RE: Follow up to previous discussion
>
> Hi Elizabeth,
>
> In my meetings with Daniel I found that the discrepancies between our CVs were due to Daniel calculating CV based on the median value of each precision run, while I was calculating CV of the entire data set for each level. When I asked him why we do this, he said that it was a way to average out the noise. I was under the impression that the coefficient of variation was meant to be, at least in part, a measure of how much noise exists in the data. By averaging out this noise before CV is calculated, the CV as a metric of assay performance becomes less meaningful. And because our calculations of CV are based on median rather than mean, this means that 2/3 of our data is entirely ignored both when calculating CV and acquiring a patient result.

Calculating the median does not ignore any data. It is statistical metric taking into account all the data. For one assay, we found the median to perform better than the mean. All other assays use the mean across the different tips. But that aside, Tyler still just does not grasp the meaning of the CV. There is no reason to report the CV of each individual tip for our assay. We happen to be running 6-tips (or six replicates) inside the device, but this what goes on inside the device really should be black box to the end user. The final reported result is all that matters. It is important for us (ie, R&D) to know tip to tip variance, but not relevant in terms of how we quantify assay precision (CV). I went over this in detail with Tyler before, as he was calculating this incorrectly before, but he still seems set this incorrect interpretation.

While I understand that calculating CV based on the medians is relevant for comparing our system to systems of our competitors, the fact that the CV of our cutoff level for Syphilis RPR drops from 43% to <20% by moving from CV of the entire dataset to CV of the medians tells me that a significant portion of our data is just noise. I believe that we should set two standards of CV that must be met in order for an assay to pass precision testing; a standard for the medians of each run, and a standard for each level's dataset as a whole.

Again, the variance across tips is not relevant - the variance for the reported value is what is quantified to assess assay performance.

Daniel also told me that for qualitative assays such as Syphilis RPR, the CV as metric of assay performance is less important than it would be for quantitative assays. I agree with him, at the end of the day the only thing that's important is delivering the correct result to our patients. However, given the high variation in our dataset, it is not surprising that when using a strict antibody index cutoff value of 1, our sensitivity was only 65% the first time we tested clinical samples and 80% the second time. The first issue I have with this is that there is no penalty for repeating an experiment. We repeat and delete rather than repeat and add. In our validation reports there is never any mention of how many attempts of precision or comparability testing it took to get the data that's presented. The second problem that I have is that our equivocal zone is adjusted and widened until we see the sensitivity and specificity that we want to report. Almost regardless of what the data looks like, we can adjust this zone until we get the 95% sensitivity that we want to see. Tellingly, out of the 247 patients that we tested, 66 of whom were Syphilis positive, more patients fell into our equivocal zone than we correctly diagnosed as being positive for Syphilis.

He makes it sounds like something inappropriate is being done, which it is not. Equivocal zones are commonly used, and expected in such qualitative assays. The approach being used for setting ours was based on common techniques. That being said, I do that our equivocal range is wider than I would like. The impact is that more patients will need confirmatory testing.

I don't know that any study was simply repeated with the original data being ignored. There have been times when the initial data was not good enough. The chemistry team then was asked to identify the root cause and make a change. And then the study was repeated.

I then asked Daniel if he thought our Syphilis test was truly the most accurate and most precise Syphilis test on the market. He said that Theranos does not claim to have the most accurate or precise tests, and that if I could find any marketing materials that make such claims that I should forward them to him. A quick google search yields a handful of articles that explicitly make these claims. Daniel agreed that the authors make sweeping statements about our assay performances, but noted that Theranos never directly made any of these claims. If well-established institutions such as the Wall Street Journal have published misinformation about Theranos, it seems it would be in our best long-term interest to correct this information in order to uphold our image of bringing transparency to blood testing.

I did note to Tyler that Theranos will have a superior product by controlling/monitoring/reducing variance across our country-wide infrastructure. This will enable us to track/trend tests results for patients in a much more robust manner compared to what is available now to patients. Moreover, I explained that the precision studies done by other companies are typically limited in terms of well controlled experiments across a few sites. But in practice, across different labs, performance is known to be much worse. I will attach below the media excerpts that he had sent me.

I then thought back to our previous discussion when I asked about our claim of having <10% CV for our assays. We checked the Theranos website together and found that we only make this claim for Vitamin D. I checked the 2-Tip validation data (we were running 2-tip protocol at the time) and found that the CVs for our three levels were 18%, 16%, and 19% when calculated based on the median of each precision run and 23%, 23%, and 25% when calculated based on the entire dataset. Here are scatter plots of the results from VitD precision testing, they don't seem to meet the standard we claim on our website for Vitamin D.

<image001.png>
<image002.png>
<image003.png>

Median was not used here – averages are.

For a while I've been giving our assays the benefit of the doubt until we see how the new 6-Tip method performs. Here is a comparison of the 7 assays we run on Theranos devices to their predicate methods. While we are now performing better than we were with the 2-Tip method, you can see that of the 7 assays we run on the Theranos system, there is only one level from one assay that shows less variation than our competitor's technology.

His Theranos TSH precision numbers do not match the validation reports. I am not sure if this is because he is calculating them himself, and using median, etc. That being said, our precision is still not as good as the Immulite for TSH. This is not new – we have discussed going back to some of these assays (such as TSH and PSA) for which we would like to increase performance, and create a next generation Theranos assay.

Immulite 3rd generation TSH	
level (uIU/ml)	total CV
0.016	12.5%
0.32	5.3%
1.3	4.6%
3.3	4.8%
7.3	5.1%
19	4.5%
39	6.4%

Theranos TSH		
	6-Tip	
Level (uIU/ml)	**CV whole dat**	**CV medians**
0.02	42.9%	34.1%
2	24.6%	17.9%
20	27.7%	20.8%

Again his number are off. Our precision numbers are instead 8.4%, 3.5% and 4.6%. This is pretty close to the predicate and better in some cases.

Immulite fT4	
Level	CV total
0.51	10.2%
0.85	7.1%
1.13	6.4%
1.49	6.0%
2.91	3.6%

Theranos fT4		
	6-Tip	
Inter mean	**whole dat cv**	**CV medians**
1.63	28.8%	14.5%
5.42	11.0%	4.0%
6.68	5.2%	3.9%

4.82	3.6%	

Calculations in the report show CVs values of 19.2%, 9.2% and 7.6%.

Immulite TT4		Theranos TT4		
level	CV total		6-Tip	
		Level	CV whole Dat	CV medians
1.8	11.7%			
2.6	10.8%	1.91	16.0%	13.9%
5.2	8.5%	3.37	16.0%	14.0%
7	6.1%	15.8	18.3%	14.6%
8.2	5.6%			
13	6.0%			
16	5.6%			

CVs in the report are actually slightly higher than what Tyler has indicated (14.7%, 13.3%, and 12.8%).

Immulite tPSA	Theranos tPSA		
"<4.6% for 3 levels of controls"		6-Tip	
	Level	CV whole Dat	CV medians
	1.4 (ng/ml)	33.8%	13.0%
	3.37 (ng/ml)	17.1%	10.8%
	10.2 (ng/ml)	24.1%	11.8%

Values in the report are 12.4%, 9.4%, and 7.3%.

Diasorin VitD		Theranos VitD
Level	CV	6-Tip

7.2	5.5%
14.7	4.2%
21.7	4.0%
35	2.9%
73	3.2%
62.7	3.1%
93.6	3.2%
115	4.2%
128	4.8%

Level	CV whole Dat	CV medians
11.7 (ng/ml)	18.6%	12.5%
28.7 (ng/ml)	19.1%	9.5%
73.6 (ng/ml)	12.1%	9.8%

Oraquick HCV	
Sensitivity	99%
Specificity	100%

Theranos HCV	
Sensitivity	99%
Specificity	94%

Values from the report: 7.5%, 6.2%, and 9.3%. Definitely on par with the reported Immulite values.

Immulite TST	
Level	Total CV
27.1 ng/dL	24.3%
86.1 ng/dL	13.0%
152 ng/dL	10.3%
280 ng/dL	9.1%
414 ng/dL	8.2%
991 ng/dL	7.2%

Theranos TST

6-Tip

Level	CV whole Dat	CV medians
90 ng/dL	19.4%	11.6%
300 ng/dL	12.5%	5.1%
1,000 ng/dL	17.4%	13.0%

Furthermore, Theranos has an inherent advantage in these comparisons due to the way we run our precision testing. While our competitors conduct their precision testing over 20 days, we do ours in 5. Accordingly, we can see that our precision

experiments are not indicative of longer-term assay performance once we begin running patient samples; our Daily Quality Control failure rate is far greater than would be predicted by our QC reference range calculations, and our internal comparison of Theranos results in proficiency testing yielded less than satisfying results. I am not sure if this analysis has been done, but we should examine our Daily QC results as if it were a prolonged precision experiment to more accurately evaluate long-term assay performance.

We did conduct all our precision runs in 5 days rather than 20. We did spread it over multiple shifts and operators per day. "20 days" is typical for FDA 510K studies, and for LDTs. Any impact of reagent stability would be reduced by compressing the study to fewer days.
If he is referring to what CLIA did last month by splitting the PT sample and running it on Theranos and on the predicate – I explained that this was not the correct procedure, and the results would not be meaningful. Moreover, that the correact procedure of doing AAP was being testing and rolled out. I do not know if he is referring to the recent AAP results for fT4, TSH, and Vit D.

I am sorry if this email sounds attacking in any way, I do not intend it to be, I just feel a responsibility to you to tell you what I see so we can work towards solutions. I am invested in this company's long-term vision, and am worried that some of our current practices will prevent us from reaching our bigger goals. I'm sorry I wasn't able to catch you for a conversation, I know how busy you are, but if you would like to discuss anything I've mentioned in person, I would be more than happy to do so.

Thanks,

Tyler

Below is the email Tyler sent me with the are media excerpts for which he expressed concerned. Note again that he expressed confusion about how we will be doing PT. I explained our PT and AAP to Tyler at some point, and that what CLIA did for that PT event was not correct. By the way, just last week, I again had to remind Mark and Adam why we are doing the AAP program like I proposed, and why we cannot split samples like they did, and why testing the PT samples on the predicate rather than on our LDTs makes the most sense.

As for finding places where Theranos' test performances are discussed, I am realizing that Theranos does not directly make any public claims that our tests are more accurate, but the authors of articles about Theranos are. The only place I can find Theranos making a claim about test performance is in this banner on our website:
<image004.jpg>
And although it specifically identifies that VitD has <10% CV, there also seems to be an implication that our other tests follow suit.

Here are some other places where it is claimed that Theranos is more precise and more accurate than current methods.
From "This Woman Invented a Way to Run 30 Lab Tests on Only One Drop of Blood"
http://www.wired.com/wiredscience/2014/02/elizabeth-holmes-theranos/
"The results are faster, more accurate, and far cheaper than conventional methods"

From "Small, Fast and Cheap, Theranos Is the Poster Child of Med Tech — and It's in Walgreen's"
http://singularityhub.com/2013/11/18/small-fast-and-cheap-theranos-is-the-poster-child-of-med-tech-and-its-in-walgreens/
"But perhaps lab tests can be made faster, easier and more accurate with a turn-of-the-last-century technology: automation. That's the bet the Silicon Valley company Theranos is making…"
"All of the diagnostic technology is integrated, which increases precision"

From "Secretive Theranos emerging (partly) from shadows"
http://www.bizjournals.com/sanfrancisco/blog/biotech/2013/09/theranos-elizabeth-holmes-walgreens.html
"But the story is scientifically appealing as well because it involves miniaturized technology — microneedles, nanotubes and other teeny-weeny stuff — that could provide more accurate medical information than that collected from traditional blood tests."

From "Breakout in Healthcare: Part I"
http://www.gingrichproductions.com/2013/10/breakout-in-healthcare/
"Theranos has developed technology that can perform all of these tests much more accurately than current laboratories, and with just a few drops of blood…Even more importantly than the greater precision and speed, Theranos has promised to deliver each of its tests for *less than half* the Medicare rate…That is a huge reduction in spending without hurting anybody except Theranos's slower, more expensive, and less precise competitors"

From the Wall Street Journal article "Elizabeth Holmes: The Breakthrough of Instant Diagnosis"
"Theranos's processes are faster, cheaper and more accurate than the conventional methods and require only microscopic blood volumes, not vial after vial of the stuff."
"Another Theranos advance is its testing's accuracy."

In regards to PT, I think most of the confusion about the process is in regards to the legality of how we conducted our PT. The Public Health Service Acts says that "the laboratory agrees to treat proficiency testing samples in the same manner as it treats materials derived from the human body referred to it for laboratory examinations or other procedures in the ordinary course of business". Which it seems we did not do, as patient samples are run on the Theranos system while the PT data we sent out were run on systems like Advia, DiaSorin, Immulite. It additionally prohibits the splitting of a sample to run confirmatory tests in another lab. While we did not necessarily send our samples out to other labs, we did split the samples to run the tests on two different sets of laboratory equipment.

Thanks,

Tyler

From: Elizabeth Holmes
Sent: Thursday, April 10, 2014 4:27 PM
To: Tyler Shultz
Subject: RE: Follow up to previous discussion

Tyler: I'm tied up with people onsite – shoot me an email with anything you wanted to cover so I can be sure it gets addressed,
Elizabeth

From: Tyler Shultz
Sent: Thursday, April 10, 2014 3:24 PM
To: Elizabeth Holmes
Subject: Follow up to previous discussion

Hi Elizabeth,

When you have time could I possibly have half an hour to follow up on our previous meeting about the RPR test? I know you are extremely busy, so I wouldn't mind waiting until an evening after the craziness of the work day dies down.

Thanks,

Tyler