November 21, 2014

Authors Response to Decision Letter for JHM-14-0296

Dear Editor:

Thank you very much for your comments as well as the outstanding suggestions and comments made by the reviewers. We have diligently reviewed each comment and have addressed the concerns and suggestions made. We have made several major changes, including restructuring the introduction to focus on performance metrics and pay-for-performance instead of VTE events, providing additional granularity for the data as suggested by the reviewers, and noting the current literature in our discussion which suggests a more conservative approach for VTE prophylaxis and further study of impact on outcomes. Below you will find detailed responses to each individual reviewer's comments. We are excited for the potential to have this manuscript published by the *Journal of Hospital Medicine*.

 Best Regards,

 Henry J. Michtalik, MD, MPH, MHS

 Division of General Internal Medicine

 1830 East Monument Street, Suite 8017 Baltimore, MD 21287, USA.

 Phone: (443) 287-8528

 Fax: (410) 502-0923

 E-mail: hmichta1@jhmi.edu

**Editor comments:**

*- Your outcome measure was compliance with risk-appropriate VTE prophylaxis. How accurate is this metric? For example, what if the patient/family refused pharm/mech VTE prophylaxis ... patient preference does not appear to be included in the contraindications in your order set. Would choosing not to order VTEp for this patient be considered "non-compliant"? If so, then I would de-emphasize the aspects of your manuscript that call for "100% compliance"*

Thank you for this insight. Our VTE prophylaxis metric is a process metric determined by the VTE risk assessment and subsequent ordering of VTE prophylaxis. In the scenario you suggest where the patient preference is not to receive prophylaxis, the assessment and order would be completed, but the patient may ultimately refuse prophylaxis. This refusal would then be documented either on the Medication Administration Record or in the Nursing Notes. If the provider should order prophylaxis and the patient refuse, the provider would still be considered compliant because an assessment and appropriate order was made. If instead no assessment was made or no prophylaxis written for (where appropriate), then the provider would be considered non-compliant. By setting the process this way, we mandate a VTE risk assessment and discussion but also respect patient preferences. We have explicitly included this scenario in our discussion: “Also, VTE prophylaxis may be ordered but intervening events, such as procedures and changes in risk status or patient refusal may prevent patients from receiving appropriate prophylaxis.31,32”

*- As the reviewers pointed out, it would be helpful to see what practices changed over time.*

We noted the comments made by the Editor and reviewers to provide granularity as to whether the changes in VTE prophylaxis were from decreased rates of not providing prophylaxis versus increased rates of appropriate prophylaxis. We agree this is an important detail and have added this information. We include these new data in the Results section: “Both inappropriate prophylaxis and lack of prophylaxis when indicated resulted in a non-compliance rating. During the three periods, inappropriate prophylaxis decreased from 7.9% to 6.2% to 2.6% during the baseline, dashboard only, and subsequent pay-for-performance periods respectively. Similarly, lack of prophylaxis when indicated decreased from 6.1% to 3.2% to 3.1% during the baseline, dashboard only, and subsequent pay-for-performance periods respectively.”

*- You may keep Figure 2 in the paper but please respond to the reviewers concerns that the risk algorithm may not fully reflect clinical practice*

We thank the Editors and reviewers for their attention to the details of our algorithm. This algorithm was instituted under the guidelines of the current time. We recognize that VTE prophylaxis recommendations have changed and the risk algorithm may not fully reflect current clinical practice. As such, we have explicitly noted this in our discussion: “Although VTE prophylaxis is recommended for patients with VTE risk factors, there are conflicting findings about whether prophylaxis prevents VTE events in lower risk patients and current studies suggest that most patients with VTE events are severely ill and develop VTE despite receiving prophylaxis.27-29”

We have also added the additional references suggested by Reviewer 1, including concerns for surveillance bias: “Similarly, hospitals with higher quality scores have higher VTE prophylaxis rates but worse risk-adjusted VTE rates which may result from increased surveillance for VTE, suggesting surveillance bias limits the usefulness of the VTE quality measure.33,34 Nevertheless, VTE prophylaxis remains a publicly reported Core Measure tied to financial incentives.4,5” As noted by Reviewer 1, VTE prophylaxis rates are a process measure often used to gauge hospital quality and hence remains a good target for our intervention.

*- You mention that the trends were the same for the 19 providers present during all three time periods. Was the converse true? (eg, were short-duration providers less likely to respond to the dashboard/incentive program?)*

We recognize that provider turnover is expected to occur in a hospitalist program and can impact trends. We presented data on the providers present during all 3 time periods to assess the robustness of the findings. We do not specifically report the impact of the program on the short-duration providers because of the heterogeneity of that group, i.e. present during different time periods, different percent efforts, and varying duration (exposure). We consequently did not include an analysis of the short-duration providers because of their heterogeneity of characteristics.

**Reviewer's Comments:**

**Reviewing: 1**

*This is a paper by Michtalik et al that asks whether the use of an individualized hospitalist dashboard +/- P4P program can lead to increases in risk-appropriate (e.g. those who could benefit get it, those who wouldn't or would be at risk of complications don't) VTE Prophylaxis for patients on the uncovered hospitalist service. Using a pre-post design with concurrent unintervened controls--patients on the teaching service where residents make these decisions with attending supervision--they find an increase of ~8% total, from 86%-->90% (with the dashboard)-->94% (with weak P4P, mean $600 total over the year).*

*So long as this paper is about the intervention and its ability to move behavior, I think it is worth publishing, but would like to see a lot more said about the marker and a little less about VTE Prophylaxis itself especially since there is a lot of recent literature that (1) this problem is not nearly as common as it was assumed to be (~2% of hospitalizations even at the high end of estimation (see e.g.* [*http://www.ncbi.nlm.nih.gov/pubmed/24100354*](http://www.ncbi.nlm.nih.gov/pubmed/24100354)*,* [*http://www.ncbi.nlm.nih.gov/pubmed/24497463*](http://www.ncbi.nlm.nih.gov/pubmed/24497463)*), and (2) hospitals that achieve very high levels of VTEP, and those achieving almost no VTEP, have virtually identical rates of post-surgical VTE (*[*http://www.ncbi.nlm.nih.gov/pubmed/24500768*](http://www.ncbi.nlm.nih.gov/pubmed/24500768)*). To me the interesting part is really the dashboard and the incentive.*

We thank the reviewer for this insight. We recognize that VTE prophylaxis guidelines have changed and there is a focus on the impact of prophylaxis on outcomes and its cost-effectiveness. We acknowledge this and have included the 3 additional citations suggested by the reviewer in our discussion.

We have cited the issue of surveillance bias in our discussion: “Similarly, hospitals with higher quality scores have higher VTE prophylaxis rates but worse risk-adjusted VTE rates which may result from increased surveillance for VTE, suggesting surveillance bias limits the usefulness of the VTE quality measure.33,34” We have also noted that most VTE events occur despite appropriate prophylaxis and included the suggested citations: “… current studies suggest that most patients with VTE events are severely ill and develop VTE despite receiving prophylaxis.27-29”

*1. INTRO: Please make this more about the literature on public reporting and dashboards and make only 1-2 sentences about VTE prophylaxis rates, mainly as a process measure often used to gauge hospital quality and hence a good target for the intervention.*

As suggested by the reviewer, we have reformatted our introduction. Specifically, we begin our introduction with healthcare reform, value-based care, metrics, and pay-for performance. We then briefly describe VTE prophylaxis interventions and individual provider dashboards. We have removed the discussion on VTE prophylaxis and outcomes.

*2. METHODS:*

*a. I would have liked to have seen a little more even about the intervention: where does it display, how many times per day, is there engagement with it as assessed by clicks, and so on. In the text it is noted that providers have access to it, but are they actually accessing it and/or how often/to what extent is it PUSHED to the provider.*

As noted by the reviewer, the dashboard is a “pull”, rather than “push”, method. Unfortunately, we were unable to record the number of times individual dashboards were accessed and cannot comment on this aspect. To strengthen this “pull” technique, we coupled it with hospitalist meetings and service reviews as noted in the methods section. We recognize this as a limitation and have included it in the limitations section: “Although we did not track individual physician views of the dashboard, we reinforced trends, deviations, and expectations at regularly scheduled meetings and provided feedback and patient-level data to individual providers.” We are currently considering developing “push” dashboards, but note that as with many other automatic prompts, engagement decreases over time. Therefore, coupling any intervention with in-person feedback strengthens the impact of the intervention, whether “push” or “pull”.

*b. The authors very appropriately note that risk-appropriate VTE rates were assessed such that getting VTEP when you are bleeding was considered "not appropriate".*

 Thank you for reviewing our algorithm.

c. I'm a bit confused however about the "automatic electronic query". This needn't be described in text, but even as a figure/box it would be nice to see what this query entails. Is it just based on what the doctor said, e.g. "this patient has or does not have contraindications" or "this patient is high risk?" Is there some other system doing this validating beyond what the provider says? It would be nice to see this spelled out since it is so central to the measured outcomes.

We apologize for the confusion. There is an automated electronic query which compares the physician ordering to the mandatory risk assessment. A specific algorithm was developed which determines the level of risk, appropriate VTE prophylaxis, and definition of compliance/non-compliance. This has been published elsewhere. Because of space limitations, we have not included it in this manuscript, but instead cite the original relevant publications: “This algorithm is published in detail elsewhere and has been shown to not increase major bleeding episodes.12,15”

*d. The P4P is very clearly described and appreciated.*

 Thank you for this comment. We provided detail on the P4P to allow other institutions to consider and modify its use, as appropriate.

*3. RESULTS:*

*a. I would have liked to see a bit more granularity around what % of patients are getting VTEP that should be and what % are getting VTEP that should not be? Combining it in this way is a bit confusing, even if it is combined in the metric itself.*

We thank the reviewer for pointing out this important difference of whether there was a decrease in non-compliance from inappropriate prophylaxis versus an increase in compliance. We now include this granularity in our results: “Both inappropriate prophylaxis and lack of prophylaxis when indicated resulted in a non-compliance rating. During the three periods, inappropriate prophylaxis decreased from 7.9% to 6.2% to 2.6% during the baseline, dashboard only, and subsequent pay-for-performance periods respectively. Similarly, lack of prophylaxis when indicated decreased from 6.1% to 3.2% to 3.1% during the baseline, dashboard only, and subsequent pay-for-performance periods respectively.”

b. If you have data about the rate of VTE and bleeding during this time, I think readers would like to see this.

We fully agree that outcome events are of great interest, particularly given recent questions about the efficacy of prophylaxis in medical inpatients. However, because of limitations of the electronic medical record, attribution of bleeding (e.g. prophylaxis versus comorbidities), and because of the limited scope of this project, no bleeding rates were collected. However, when the initial algorithm was published, this was assessed and there was no increase in bleeding events (References 12 and 15). We also included this statement in our discussion section: “Our study was underpowered to detect these potential differences in VTE rates and although the algorithm has been shown to not increase bleeding rates, we did not measure bleeding rates during this study.12,15”

Our study was also underpowered to detect differences in VTE rates. For the reviewer, with respect to the VTE rates, we were able to electronically extract this data using the AHRQ Patient Safety Indicator. The 13 VTE events were as follows: 8 events during the baseline period (all on appropriate prophylaxis); 1 event during the dashboard only period (not on appropriate prophylaxis); and 4 events during the subsequent pay-for-performance period (all on appropriate prophylaxis). Unfortunately, because of the rarity of these events, we cannot definitively comment on the direct impact of this program on VTE rates. We did, however, include the following statement in the Discussion: “Our institutional experience suggests that the majority of VTE events occur despite appropriate prophylaxis.30”

*4. DISCUSSION:*

*a. First para--how are metrics external motivators if peer norms are not? This was a bit unclear to me; I would just say "extrinsic = payments and intrinsic = metrics"--this seems cleaner to me.*

Thank you for bringing this ambiguity to our attention. We have clarified this by removing “metrics” from the sentence: “Our study evaluates the impact of both extrinsic (payments) and intrinsic (professionalism and peer norms) motivation.” We did not specify metrics because metrics could be tied to either payments (extrinsic) or professionalism (intrinsic).

*b. First para--the total disbursement being < incremental cost of 1 preventable VTE adds little for me and I'd take this out. $10,000 is $10,000 and it's not clear to me that going from 86% risk-appropriate VTEP to 94% risk-appropriate VTEP will prevent 1 more VTE per year (and/or not cause a bleed).*

We have removed this sentence from both the abstract and manuscript.

c. Third para--they point out the VTEP rates in the teaching svc vs. the uncovered service were different, but in the results I didn't see these explicitly statistically compared (difference in difference?)--this would be appropriate if we are truly using the other groups as a control. This could be done in the Figure itself.

We thank the reviewer for this suggestion. We have added the following to the Results section: “However, there was no statistically significant difference between the general medicine service teams and hospitalist service during the dashboard only (p=0.15) and subsequent pay-for-performance (p=0.76) periods.” We did not include the different p values on the Figure to prevent confusion from whether these p values represented changes in slope during each time period or a comparison to the general medical services.

*d. Limitations para--would include the fact that VTEP as a measure is complex and prone to dispute, and also the fact that actual VTE and bleeding rates are not included here. Also that the hospitalist work force isn't stable over time (lots of functional PGY-4s and PGY-5s, so in the later periods the successful residents from the earlier periods might have aged in, and this might have mattered as much as P4P etc) and then explicitly make the point that the subanalysis addressed this.*

We thank you for these suggestions. In our limitations paragraph, our second limitation notes that VTE prophylaxis measures are complex and prone to dispute: “Second, it is difficult to show actual changes in VTE events over time with appropriate prophylaxis. Although VTE prophylaxis is recommended for patients with VTE risk factors, there are conflicting findings about whether prophylaxis prevents VTE events in lower risk patients and current studies suggest that most patients with VTE events are severely ill and develop VTE despite receiving prophylaxis.27-29 Our study was underpowered to detect these potential differences in VTE rates …”

In the methods section, we explicitly note that “this algorithm is published in detail elsewhere and has been shown to not increase major bleeding episodes.12,15” In the limitations section, we have also now included that bleeding rates were not measured: “although the algorithm has been shown to not increase bleeding rates, we did not measure bleeding rates during this study.”

Similarly, we have explicitly noted the reviewer’s limitation of residents potentially aging into the hospitalist service: “Although it was possible for successful residents to age into the hospitalist service, thereby improving rates of prophylaxis based on changes in group makeup, our subgroup analysis of the providers present throughout all phases of the study showed our results to be robust.”

*5. FIGURES: I found Figure 1 helpful and Figures 2a and 2b not that helpful. As near as I can tell this study is primarily about the dashboard and P4P, not about the order set; so I would stuff the latter two in an appendix.*

In this manuscript, we are balancing succinctness and detail. As per the Editor recommendations, we have left Figure 2 in the manuscript rather than shift it to an appendix. Similarly, as the Editor and reviewers have commented, the manuscript should focus on the dashboard and pay-for-performance programs so we have not provided a detailed flow diagram of the algorithm as this will detract from the main messages of the manuscript, especially given the changes in VTE prophylaxis recommendations as this reviewer has pointed out.

*Instead as Figure 2 I would include a box indicating flow of decision making around VTEP appropriateness, e.g. "Does this patient have contraindications yes/no-->if Yes = in appropriate to prescribe (n=...) if no move on to, Does this patient have indications for VTEP or is this patient low risk; if low risk = inappropriate (n=...)to prescribe if high risk appropriate to prescribe (n=...)"*

 We addressed this comment in the preceding response.

*Figure 3 I would include p values for differences, and/or whether this was underpowered. Also, the shape of the graph obscures the fact that the difference at the end of the day is between 94-95% vs. 90-91%, which is not clearly clinically meaningful. I suppose it's necessary to show that there is a difference but I'm not enthused about cutting off the axis at 75% and would rather see it start at 0%.*

For Figure 3, we stated the p values in the text: “During the dashboard only period, on average providers improved in compliance by 4% (95% CI: 3, 5; p<0.001). With the addition of the pay-for-performance program, providers improved by an additional 4% (95% CI: 3, 5; p<0.001).” We did not add the p value to Figure 3 to prevent confusion on whether this p value represented the statistically significant difference of the hospitalist group over time versus the general medicine teams.

With respect to the axis at 75%, we strived to balance graphical clarity and accuracy of comparisons. We did not choose an axis cut-off of 0 because it would be difficult to visualize the change over time. We also did not choose an axis cut-off of 90% because we believed it would visually exaggerate the impact. Therefore, we chose an intermediate cut-off of 75%.

**Reviewing: 2**

Comments to the Author:

*Overall comments:*

*Dr Michtalik and colleagues report the results of using a dashboard and pay-for-performance bonuses to improve the rate of appropriate prescribing of VTE prophylaxis by a hospitalist group. This is a relevant topic and the study is well-done and well-described. The authors conclude that the both reporting and payment were associated with improvements in appropriate VTE prophylaxis, but that the transparent reporting of individual metrics had the greatest impact on performance.*

Major:

*I think the authors should provide more information about what constitutes appropriate prescribing, and how the actual use of pharmacological, mechanical or absence of prophylaxis changed over time. Was there a change in documentation of contraindication? Or increase in prescription of prophylaxis?*

We thank the reviewer for this insightful comment. We have included both inappropriate prophylaxis and lack of prophylaxis data in the Results section as per this reviewer and Reviewer 1’s comments: “During the three periods, inappropriate prophylaxis decreased from 7.9% to 6.2% to 2.6% during the baseline, dashboard only, and subsequent pay-for-performance periods respectively. Similarly, lack of prophylaxis when indicated decreased from 6.1% to 3.2% to 3.1% during the baseline, dashboard only, and subsequent pay-for-performance periods respectively.”

*I would like to see a broader discussion about the implications of these findings for other sites/groups and other conditions cared for by hospitalists*.

In our conclusions, we provide a broader discussion and implication of these findings: “The specific goal (risk-appropriate VTE prophylaxis) and benchmarks (80%, 85%, 90%, 95%) can be individualized to a particular institution. For example, if readmission rates are above target, readmissions could be added as a dashboard metric. The specific benchmark would be determined by historical trends and administrative targets. Similarly, the overall financial incentives can be adjusted based on the financial resources available.”

*Methods – overall the methods are appropriate for longitudinal evaluation.*

*The authors should specify what constitutes appropriate prescribing more clearly. They mention that patients with contraindication should not receive prophylaxis, but how did they categorize patients without any identified risk factors? Was mechanical prophylaxis considered in the determination of appropriate care, or only based on pharmacological?*

We thank the reviewer for the attention to detail. In our methods section, we provide the references for the algorithm and risk classification: “Based on ACCP VTE prevention guidelines, risk-appropriate prophylaxis was determined using an electronic algorithm that categorized patients into risk categories based on the presence of major VTE risk factors” and cite the references. We are limited both by scope and space restrictions in this manuscript. Our algorithm did include both mechanical and pharmacological prophylaxis. We have clarified this by modifying our methods to state “the system then prompts the provider with a risk-appropriate VTE prophylaxis recommendation that the provider may subsequently order, including mechanical prophylaxis (Figure 2B).”

*Further, the authors should present how the actual use of pharmacological, mechanical or absence of prophylaxis changed over time. Specifically, was the improvement in appropriate prescribing due to the fact that more patients considered to be high risk were prescribed prophylaxis, or because more patients with contraindication did not receive pharmacologic prophylaxis? Could it be that the incentives changed documentation more than actual delivery of prophylaxis?*

This comment is similar to the one provided by Reviewer 1 and addressed with additional data presented in the Results section: “Both inappropriate prophylaxis and lack of prophylaxis when indicated resulted in a non-compliance rating. During the three periods, inappropriate prophylaxis decreased from 7.9% to 6.2% to 2.6% during the baseline, dashboard only, and subsequent pay-for-performance periods respectively. Similarly, lack of prophylaxis when indicated decreased from 6.1% to 3.2% to 3.1% during the baseline, dashboard only, and subsequent pay-for-performance periods respectively.”

*The authors appear to use the physician documentation of risk/contraindications as the gold standard to determine appropriateness. This seems reasonable but would be interesting to note if this has this been validated in their system.*

We thank the reviewer for this observation. We have not validated the physician documentation of risk/contraindication. We expect the medical record to accurately reflect risk factors.

*The authors mention that the dashboard is on the institutional portal – is there any way to track views of this site by hospitalists to indicate usage?*

This comment is similar to that of Reviewer 1 and previously addressed. Our intervention was a “pull” intervention. We did not track individual views by hospitalists. We explicitly added this limitation to our conclusions: “Although we did not track individual physician views of the dashboard, we reinforced trends, deviations, and expectations at regularly scheduled meetings and provided feedback and patient-level data to individual providers.”

*Were the individual results blinded to others*?

Providers could only access the aggregated and his/her individualized dashboards. We clarified this in the manuscript: “Each provider was able to access the aggregated dashboard (showing the group mean) and his/her individualized dashboard using an individualized login and password for the institutional portal.”

*Was the P4P program only applied to VTE proph or also the other metrics listed in dashboard?*

The P4P Program was applied to other metrics listed in the dashboard but the other metrics are beyond the scope of this manuscript.

*The authors could be more clear about the physician group description.*

 *1) It becomes apparent later in paper, but it would be helpful if authors indicate clearly that the hospitalist group does not cover teaching services.*

 *2) The authors should clarify whether there is any overlap between the hospitalist under study and the physicians that provide care on the other teaching services.*

 *3) the authors state that the group covers one 17-bed general medicine unit. It would be helpful to understand the other services that the group covers, particularly as compared to the other teaching services used as a contemporaneous control.*

We thank the reviewer for pointing out these ambiguities. We have addressed this by detailing the services of the hospitalist group in one section: “The scope of diagnoses and acuity of patients admitted to the hospitalist service is similar to the housestaff services. Some hospitalist faculty serve both as hospitalist and non-hospitalist general medicine service team attendings, but the comparison groups were staffed by hospitalists for less than 20% of the time.”

We also acknowledge that there may have been a cross-contamination effect of hospitalist faculty who could serve as both hospitalist and non-hospitalist general medicine service team attendings. This, however, would attenuate any impact of the P4P program and thus our effects may in fact be greater than reported. We have added these statements to the discussion section in the limitation paragraph: “Similarly, there may have been a cross-contamination effect of hospitalist faculty who attended on both hospitalist and non-hospitalist general medicine service teams. This, however, would attenuate any impact of the programs and thus the effects may in fact be greater than reported.”

*Results*

*The results are clearly stated.*

*As above, I think there is value in showing the proportion of patients who received pharmacologic therapy over time in the hospitalist group and the teaching group. If there are no differences over time or between groups could just mention that clearly in text.*

We have addressed this similar comment from Reviewer 1 by explicitly commenting in our manuscript on the measure of non-compliance, separated by not prescribing prophylaxis versus ordering inappropriate prophylaxis, and make a comparison to the teaching group.

*Discussion/Conclusions*

*Overall the discussion is well-written and clear, and not repetitive.*

*I think it could benefit by a clearer discussion of the components that are necessary to apply this methodology at other locations and with other conditions. The 4th and 6th limitations are very instructive and could form the basis of this discussion is the authors choose to modify their approach.*

We thank the reviewer for this comment. Our discussion section focuses on the rationale for the study design, generalizability, and potential implications. We recognize the resources needed to create the dashboard, its continued maintenance and monitoring of performance, and the payment of financial incentives. We are limited by space restrictions, but have expanded our final paragraph to include an example such as readmissions: “The specific goal (risk-appropriate VTE prophylaxis) and benchmarks (80%, 85%, 90%, 95%) can be individualized to a particular institution. For example, if readmission rates are above target, readmissions could be added as a dashboard metric. The specific benchmark would be determined by historical trends and administrative targets. Similarly, the overall financial incentives could be adjusted based on the financial resources available. Other process measures, such as influenza vaccination screening and administration, could also be targeted.”

*The authors mention that one limitation is that it is a single-site study – I think though that they may underestimate the strength of data support for this initiative, including a clear decision algorithm in CPOE. This could be made more clear, particularly as it is relevant for generalizability.*

We appreciate this clarification and have rewritten this section to state: “There was strong data support and a defined CPOE algorithm for this initiative. Multi-site studies will need to overcome the additional challenges of varying service structures, and electronic medical record and physician order entry systems.”

*It would be helpful to give insight into the generalizability of these findings to other conditions – specifically, the authors should provide some insight into how this knowledge can/should be translated to the complicated conditions cared for by hospitalists. What are limitations, particularly in areas without clear cut evidence? Though this may go beyond scope of paper, the authors could consider referencing how they have chosen to handle the other metrics included in the dashboard.*

With respect to other potential conditions/priorities for which these study concepts could be applied, we expanded our last paragraph to include a discussion of readmissions. We also noted that “other process measures, such as influenza vaccination screening and administration, could also be targeted.”

In the Methods section for the hospitalist dashboard, we specify that “hospitalist program leaders met with hospital administrators to create a hospitalist dashboard which would provide regularly updated summaries of performance measures for individual hospitalists.” The measures were selected based on institutional priorities and financial incentives assigned according to hospital priority and funds available. Providing detailed discussion of the additional metrics is unfortunately beyond the scope of this manuscript.

*The authors indicate that the timing of the financial incentive may impact the program’s effectiveness. It also seems like the amount is critical. How was the size of the incentive calculated – is there reason to think that this is an “appropriate” amount?*

The financial incentive was determined according to hospital priority and funds available. We have added this statement to our Methods section. If there was no change in compliance in our study, then one potential cause could have been an inadequate incentive size. However, to address the specific impact of incentive size, a separate study would need to be done to determine a threshold for behavioral change which would subsequently need to be validated in other settings. This is an interesting area of research, but cannot be addressed by our study design.

**Reviewing: 3**

*Comments to the Author:*

*Biggest thoughts on the paper:*

*First, this is well written and relevant for the JHM readership, and suitable for publication. I have two major thoughts and some minor ones for the draft:*

*We need information on VTE rates or a discussion of how the absence of this data makes the result more about influencing behavior rather than influencing outcomes. This is particularly important because if I understand the prophylaxis algorithm from reference 15 correctly (note that almost no readers will track that down), you were really pushing people toward universal prophylaxis which was not standard of care and has become even less favored since the new guidelines came out.*

We thank the reviewer for these insights. We have provided the VTE rates in our response to Reviewer 1 but did not include them in the manuscript because our study did not specifically collect VTE rates nor was powered to detect differences. We have included the additional references noted by Reviewer 1, which suggest that improved prophylaxis may not actually impact outcomes. Unfortunately, because of the rarity of these events, we cannot definitively comment on the direct impact of our program on VTE rates. We did, however, include the following statement in the Discussion: “Our institutional experience suggests that the majority of VTE events occur despite appropriate prophylaxis.30”

*I also think it should be explicitly acknowledged that on fig 3 improvement seems to begin before intervention raising the question of whether another factor influenced behavior. This is the inherent limitation in time trends—if you look at Obama as an economic intervention and compared two time periods, the numbers would suggest he caused an economic mess, but he actually inherited one.*

We explicitly recognize this limitation in our limitations section: “Third, there may be an unmeasured factor specific to the hospitalist program which could potentially account for an overall improvement in quality of care. Although the rate of increase in appropriate prophylaxis was not statistically significant during the baseline period, there did appear to be some improvement in prophylaxis towards the end of the period.”

*Specific Comments:*

*Abstract:*

*“academichospitalists.” this is one of several fused words*

This typographical error did not appear in our uploaded manuscript file so appears to be due to a technical glitch; we hope that these fused words will not be a problem in the published version.

*Introduction:*

*“embolismeach.” Consider rephrasing because it sounds like DVT might be a mortal illness, or saying that prophylaxis or prophy rates are a performance measure rather than VTE itself.*

In restructuring the introduction, these statements were removed.

*Methods:*

*“systemthen.”*

 This typographical error did not appear in our uploaded manuscript file.

*It’s not clear to me how the prophy recommendation works. In Fig 2B, there is apparently a blank under recommended prophy and then there are several options. Is the blank usually populated with a specific recommendation based on what the provider entered? If so, how is it generated—how does the algorithm decide whether someone should SCDs, TEDs, or a foot pump, since there are three mech options?*

We did not use a screen shot with a specific recommendation since this might have caused confusion. For a recommendation to be shown, we would have had to select specific risk factors to cause the recommendation field to populate. We, however, describe both in the text and figure legend that once the risk factors are selected, a recommendation is populated. The algorithm itself is cited in the methods section: “Based on ACCP VTE prevention guidelines, risk-appropriate prophylaxis was determined using an electronic algorithm that categorized patients into risk categories based on the presence of major VTE risk factors (Figure 2A).12,15,20” Because of space constraints, we did not specifically include this prior published algorithm.

*I was interested in finding out how the system dealt with the Padua scoring system recommended by ACCP 9, and how an ortho patient on medicine would be dealt with, but looking at the study start time / references, I realized we are dealing with ACCP 8 recommendations. I think that’s worth clarifying because the landscape for VTE prophy has changed at lot (or at least the guidelines have; don’t think many people follow them). Perhaps in the discussion there’s a spot for a line about the change*.

We thank the reviewer for these astute observations. We recognize that the landscape of VTE prophylaxis is changing. Based on this comment and others, we have specifically noted these challenges in the limitations section of our conclusion: “Second, it is difficult to show actual changes in VTE events over time with appropriate prophylaxis. Although VTE prophylaxis is recommended for patients with VTE risk factors, there are conflicting findings about whether prophylaxis prevents VTE events in lower risk patients and current studies suggest that most patients with VTE events are severely ill and develop VTE despite receiving prophylaxis.27-29”

*Following your reference (Zeidan, 15) to your protocol, I was surprised to find out that it appears to represent a DEFAULT position of using prophylaxis. High risk means q8 and not high risk means q12 heparin, and mech prophy is recommended for those who can’t take heparin. This is what ACCP 8 recommends:*

 *6.0.1. For acutely ill medical patients admitted to hospital with congestive heart failure or severe respiratory disease, or who are confined to bed and have one or more additional risk factors, including active cancer, previous VTE, sepsis, acute neurologic disease, or inflammatory bowel disease, we recommend thromboprophylaxis with LMWH (Grade 1A), LDUH (Grade 1A), or fondaparinux (Grade 1A). [under contraindications it says GCS or IPC not foot pumps]*

 *This is actually quite far from default prophy. You have to have serious pulm disease or CHF OR be bedbound with other risk factors. So while you say you would “ding” providers for over treatment (see below) this is actually what you’re recommending compared with the actual guideline.*

The VTE prophylaxis algorithm was developed based on guidelines at the time of implementation and recommendations from institutional experts. We provide these citations in the Methods section. We recognize that the landscape for VTE prophylaxis is changing and also cite this in our conclusion section (see preceding response). Our VTE workgroup is currently examining these changes. The main focus of this manuscript, however, is the application of performance measures to individual physicians and the potential impact of a dashboard and pay-for-performance program upon a performance metric.

*Looking at this table in detail, one of the contraindications is aPTT>1.3. That’s an APTT ratio then?*

aPTT>1.3 should reflect a ratio. We have corrected this in the legend for the figure.

*Is there no guidance or requirement for what constitutes “high risk of bleeding?” did the study attempt to capture use of these indicators because clicking vague and unverifiable contraindication buttons is a primary way to escape from protocols and would likely vary by individual.*

High risk of bleeding was a clinical determination. We have clarified this in the text in the methods section: “Both an assessment of current use of anticoagulation and a clinically high risk of bleeding were also included (Figure 2A).” The algorithm was described and validated in prior, cited manuscripts.

*Is there a reason the order was not mandatory? At my facility providers MUST pick a level of risk and MUST pick an associated management option (a medley of either a drug, device, or designation of low risk or contraindications).*

During physician engagement sessions, meetings of the VTE workgroup, and piloting of the VTE prophylaxis order set, a mandatory prophylaxis order was not implemented. We added the rationale for this in the methods section: “This allowed the physician discretion to choose amongst various pharmacological agents and mechanical mechanisms based on patient and physician preferences.”

*Were providers penalized for using judgment, eg, providing heparin sc q12 when someone had 45,000 plts or an INR of 1.6? Sounds like they would be. I can envision a hypercoagulable patient with a rising or subtherapeutic INR on Coumadin who needs either prophy or full anticoag, for example, and this would count as “current use of systemic AC” and “INR > 1.5” if the provider filled the form out honestly.*

Both the benefit and limitation of an algorithm is that it rigidly applies pre-specified rules. Providers falling outside of the recommendations would be considered non-compliant. We attempted to balance the standardization and individualization of VTE prophylaxis by including both objective data (e.g., platelet counts) and clinical judgment (e.g., high risk of bleeding).

*Might help to order the methods section starting with the order set, then the dashboard, then the payments, because these are really the sequential steps in ongoing QI. In other words, you could start with the prior effort/ current state, then add intervention one and two chronologically. Right now it goes dashboard-current state-P4P.*

We attempted to organize this manuscript in several different ways. Prior to submission, we ordered it as suggested by the reviewer but found the flow to be interrupted and the parallel structure difficult to maintain. We subsequently decided to present the overall dashboard first, then define the metric in detail, and finally, add the pay-for-performance program. We found this order to be more succinct and maintain the flow and parallel structure, but recognize that there may be different ways to handle this.

*Results*

*Why did you analyze “over 3144” admissions? Can you just give the actual number if you know it? Or did you mean you looked at 3144?*

We thank the reviewer for the clarification. We have corrected the sentence to state that 3144 admissions were analyzed.

*“4non-hospitalist”*

 This typographical error did not appear in our uploaded manuscript file.

*The authors apparently looked at the total % compliance for three periods and report that data first, but then they also looked at the slope of the best fit line. Both of these analyses use fig 3, but there’s an intervening analysis on a subgroup. This is confusing because we leave fig 3 and come back to it. At first I thought it was a different data set and they linked the wrong figure. Consider exhausting analysis of fig 3 (means and slope) then proceeding to the subgroup of consistent providers.*

We thank the reviewer for identifying this intervening text. We have restructured the results section to avoid interrupted referencing of Figure 3. This section now presents the individual provider and group improvement, followed by the rate of change in VTE prophylaxis and concludes with the subgroup analysis. We believe these changes have improved the flow of the results section.

*This section had another confusing feature: the title is Venous Thromboembolism Prophylaxis and Compliance. This implies that two separate things will be measured: rates of prophy and compliance with the protocol (different because of possible overuse?). However paragraph one and the first part of two is all about “compliance.” Then there is mention of the rate of “VTE prophylaxis improvement” (does this distinguish prophylaxis and compliance?). The last 2 sentences were confusing:*

*“The addition of the pay-for-performance program did not significantly increase the rate of compliance (0.09% per month; 95% CI: -0.57, 0.75; p=0.78). Overall percent compliance, however, continued to improve with the addition of the pay-for-performance program, but began to plateau at 95% compliance by the end of this study.”*

*This left me wondering if “compliance” and “overall percent compliance” were different things. Is the first actually the rate of improvement of compliance? I would try to be entirely consistent in what is being reported and use the fewest number of terms here. And if something “continued to improve” I would specify if this means in a statistically significant or qualitative way, because no p value is presented for the claim. Perhaps you could say the % continued to increase but the difference was not statistically significant?*

We have made significant changes to this section based on the reviewer’s comments. We have changed the title of this section to “Venous Thromboembolism Prophylaxis Compliance” to be clearer. We have also removed the values of the individual slopes for two of the three different time periods and instead only provided p values to be more succinct. We have also removed the confusing sentence and restructured the paragraph. It now reads: “We also analyzed the rate of VTE prophylaxis compliance improvement (slope) with cut-points at each time period transition (Figure 3). Risk-appropriate VTE prophylaxis during the baseline period did not exhibit significant improvement as indicated by the slope (p=0.23) (Figure 3). In contrast, during the dashboard only period VTE prophylaxis compliance significantly increased by 1.58% per month (95% CI: 0.41, 2.76; p=0.01). The addition of the pay-for-performance program, however, did not further significantly increase the rate of compliance (p=0.78).”

*The biggest result here that’s not displayed or mentioned is VTE RATES. This could be mined from your EMR, could it not? Is this not a crucial piece of data? The reader does not know if your program improved prophy and prevented VTE (at a possible cost savings) or whether you set up a program, paid out monies, and merely increased prophy rates in low risk patients who weren’t going to have VTE thus resulting in no change. The implementation of your strategy hinges on this data. If at all possible, this should be added to the results. If it is not possible, I think an expanded mention in limitations would be appropriate. This is especially because you apparently were recommending default prophy (a specific limitation because this wasn’t standard of care, and the 9th ACCP got more restrictive). It's easier to reward people for % doing something when you're aiming for 100% receipt; since current VTE prophy standards are more nuanced, there are people who should not be treated and harder to implement such a reward and penalty system.*

We thank the reviewer for this comment and have addressed this in our prior responses to Reviewers 1 and 2.

*Discussion:*

*“VTEprophylaxis”*

*This is a quibble, but usually I would say the study DID something in the past tense, and that your program may still DO something in the present tense. You really studied a program that “attributes an individual performance metric” etc, rather than had a study that “attributes…”*

We have changed this section of sentences to the past tense as suggested by the reviewer: “It specifically attributed an individual performance metric, VTE prophylaxis, to an attending physician, provided both individualized and group feedback using an electronic dashboard, and incorporated a pay-for-performance program.”

*You hypothesize that there may be an “enhanced ability to educate residents about quality improvement efforts.” Aren’t the hospitalists the faculty who drive VTE prophy rates / the source? In our institution, the hospitalists are relatively easy to reach and of the same mind on most quality issues while the residents are focused on day to day tasks, scattered and harder to reach, so this seemed somewhat odd.*

The main message of this sentence was that residents may be more willing to accept standardized order sets and automated recommendations as compared to attending physicians, rather than apply individual physician experience and preference. We have clarified this by modifying the sentence to read: “Initially the compliance of the general medicine teams with residents exceeded that of the hospitalist attending teams, which may reflect a greater willingness of resident teams to comply with order sets and automated recommendations.” We recognize that this may vary at individual institutions.

*“However, the VTE prophylaxis metric is a national Joint Commission performance standard and attending physicians are responsible for ordering appropriate VTE prophylaxis.” This may be true, but I’m not sure what this changes about the scope of your study. We still don’t know if it works on resident led services. Why not incentivize the attendings for the rates on those services too?*

We have removed this Joint Commission sentence for clarity. In the future, as suggested by the reviewer, we hope to complete expanded studies and examine the impact of this program on additional services and team structures.

*“butintervening”*

 This typographical error did not appear in our uploaded manuscript file.

*This whole sentence could be rethought: “Also, VTE prophylaxis may be ordered butintervening events, such as pharmacy delays, procedures, or patient refusal may prevent patients from receiving prophylaxis.” First, didn’t you assess orders, such that a pharmacy delay (usually negligible anyway) and refusal wouldn’t affect the rate? Second, procedures usually result in prophylaxis dropoff because an order is stopped and not restarted, rather than not ordered to begin with. So I think a clearer way of stating this is that your program doesn’t address changes in patient status: coming off and back on prophy for procedures, and changes in risk status like improvements or worsening (of what began as “obs”).*

We thank the author for this clarification. We have modified our sentence to more accurately reflect that changes in patient status and intervening procedures may alter/interrupt appropriate VTE prophylaxis: “Also, VTE prophylaxis may be ordered but intervening events, such as procedures, and changes in risk status or patient refusal may prevent patients from receiving appropriate prophylaxis.31,32”

*“Third, there may be an unmeasured factor specific to the hospitalist program which could potentially account for an overall improvement in quality of care. However, there were no other VTE-related provider feedback programs being simultaneously pursued during this study.” --well, okay; many things change practice besides other VTE related provider feedback programs, so this does not address the underlying issue that this is a time series rather than an RCT. For example, maybe there was a grand rounds on a death and under prophylaxis (education). Maybe the faculty were becoming aware of the guidelines released around the time of your baseline data collection start time. I think the key issue here is for the discussion to acknowledge that the VTE prophy rate started going up BEFORE the dashboard was implemented—about 8 months before, then there was sustained improvement that lasted 20 months. You specifically say your greatest improvement occurred during feedback only, but actually, it began before that and extended after it. If you showed your fig 3 without vertical bars and asked blinded hospitalists when QI happened they would probably draw their guesses at months 16 and 36.*

We assessed the rate of change in appropriate VTE prophylaxis by measuring the slope during each of the three time periods. During the baseline period, there was not a statistically significant slope. However, as pointed out by the reviewer, we recognize that there appeared to be an upslope towards the end of the period. We have now explicitly acknowledged this: “Although the rate of appropriate prophylaxis was not statistically significant during the baseline period, there did appear to be improvement in prophylaxis towards the end of the period.”

*“Continued provider feedback and engagement are critical for progressive success, especially to decrease variability in care at the attending physician level; financial incentives alone will not likely result in sustainable behavior changes.” --This is a little confusing because it’s very hard to imagine a program continuing to pay doctors without giving them simultaneous feedback. The payor will be keenly interested in getting value for the $ and linking it to feedback for improvement purposes. The better point (frequency of payments) was already made above.*

We thank the reviewer for making this clarification. We have removed the comment on financial incentives alone.

*References:*

*14 is missing the year so I would just once-over the formatting of all for consistency*

We have included the missing year in this reference.

*Lastly:*

*Does JHM not require a statement that this study was IRB approved or that IRB waived or that it wasn't submitted because it was a QI project? To some extent, this was a study on hospitalists providing care. We all know there was no appreciable risk, but that has never stopped an IRB from commenting!*

The VTE prophylaxis order set and data collection and analysis were approved by the Johns Hopkins Medicine IRB. The dashboard and pay-for-performance program were initiated by the Institution as part of a “proof of concept” quality improvement project. We have included these statements in the manuscript in the Methods section at the end of the pay-for-performance description.