# Peer Review File

Article information: https://dx.doi.org/10.21037/atm-2022-58

**Reviewer A**

This study provides further evidence that CYP2D6 phenotype has no clinically significant effects on oxycodone response. The results are important to disseminate in that regard. However, the paper would benefit from editing throughout to improve the clarity of the message and presentation of the data.

1) The authors mention functional phenotype throughout the paper, but no functional data are provided.

Functional phenotype referred to having at least one functional allele. However, due to the confusion and with no added benefit to studying this, we have removed this from the revised manuscript.

2) If comparisons were not statistically significant, the authors should not state that the incidence of side effects or %AUC for pain were higher or lower in one group compared to others. The only conclusion can be that there was no difference.

This has been corrected, and these statements removed from results under AUC, RD and Emesis.

3) Page 5, lines 3-6: unclear as what effects were higher or lower with oxycodone vs codeine.

We have clarified that while codeine is a prodrug and only its metabolite is active, oxycodone is already an active drug.

4) Page 5, starting on line 14: If CPIC guidelines concluded that the evidence was insufficient to make firm recommendations for oxycodone, then why would they subsequently conclude that oxycodone has the "highest potential for actionability"?

This paragraph has been edited to further clarify our justification for studying the possible association between different *CYP2D6* genotypes and oxycodone analgesic effects and side-effects. We have removed the phrase “highest potential for actionability”

5) Unclear as written what the differences were between the 2 cohorts studied. The first appears to be a research cohort, but the research methodology is not well described. It would appear that no genotype guided recommendations were provided for this cohort, but that is not really stated anywhere. Similar description is needed for the 2nd cohort. The authors should also comments on if/how any differences between the cohorts could influence results.

The cohorts have been described better. Since both cohorts were included in an observational study, we do not expect an innate bias based on the cohort.

6) Incomplete sentences in the methods and parts are unclear. The paper would benefit from a thorough review and editing.

Methods has been revised for clarity and completeness.

7) The data related to reference #51 (Thomas et. al.) are not reported correctly. Oxycodone use in that study was very low vs 72% taking the drug in the high risk group.

Reply 7: This discrepancy has been corrected in the manuscript.

Changes in the text: Page 16 lines 1 and 2

**Reviewer B**

Merchant et al. have conducted a potentially interesting study that has an underlying question which has high relevance to modern post-operative pain prescribing – that of attempting to better characterize sources of inter-individual variability in opioid effectiveness and tolerance. Specifically, the authors focus on CYP2D6 as a well-known pharmacogene which has previously shown mixed results regarding whether the drug oxycodone is influenced by CYP2D6 variability with respect to pain outcomes.

The authors themselves are well-positioned to conduct this type of study given a clear track record of interest in this field (e.g., multiple awarded patents related to genetic personalization of opioid therapy). As an example, Figure 1 (which is background on the topic) is elegant and really nicely done – one of the best such figures I have ever seen. Unfortunately, the authors' current analysis, while detailed, in its present form is plagued by multiple methodologic shortcomings, and a lack of clarity in writing in other places.

Thanks for your review and the constructive feedback.

Major critiques:

1) A major flaw is that the authors fail to account for the total amount of opioid(s) administered – which certainly could directly affect the level of pain control (Pain AUC%, the primary endpoint). A large proportion of patients, for example, switched between oxycodone and hydromorphone during the period under evaluation, and this is not accounted for in any systematic way. If administered doses are actually recorded and known, as the authors said, why don't the authors take advantage of these data and simply calculate morphine equivalents? This would be a standard way of quantifying total opioid received, which then could be used to compare different genotype groups.

Thank you for this suggestion. We have revised our outcome and calculated morphine equivalents/kg for all oral and intravenous opioids used (oral oxycodone/ hydromorphone, IV hydromorphone/morphine). Since we are mainly interested in the effects of opioids metabolized by CYP2D6, we decided to use the oxycodone/Total MEq as the primary outcome. This also takes into consideration the need for oral opioid changes, and is adjusted for pain AUC over that period.

2) The primary conclusion of the paper (first line of abstract, and elsewhere) is stated strongly, but is actually potentially misleading or even incorrect. While the authors' data as calculated do not demonstrate a difference in pain outcomes based on CYP2D6 genotypes, perhaps this is actually expected – that "no difference" would be found on pain scales because providers would adjust their prescribing in real time based on the patient's pain, to ACHIEVE pain control. In other words, reaching "adequate pain control" (or equivalence in pain control) is actually an expected and required quality metric in modern healthcare, so any patients who are not achieving pain control with a given first opioid (e.g. oxycodone) would be dose-titrated or switched to an alternative (hydromorphone). This is in fact what the authors' own data shows happened. So how can one conclude that genotypes did not affect outcomes when equivalence in pain is exactly what would be sought after? Smith et al. (Genetics in Medicine) which the authors cite (reference #29) showed this same phenomenon, in fact. The phenotype being used (pain AUC) is therefore probably the wrong primary phenotype to expect a "difference". How to account for this relates directly to Major Critique #1.

Please see our response to your prior critique. Changing the primary outcome to an opioid dose-based outcome has allowed us to avoid using pain AUC as an outcome. We agree that this outcome is superior as pain control is usually always achieved by adjusting opioid choice and dose.

3) The conflation of "efficacy" and "side effects" in the same "outcomes" 'bucket' make the entire analysis confusing. The authors should pick one as the primary analysis endpoint and concentrate the analysis on that. The other should be secondary.

In the revision, we clearly demarcated the primary and secondary outcomes or endpoints as suggested.

4) Despite all the methodologic shortcomings (which should be revised and reanalyzed as suggested above), Figure 2 seems to actually suggest that there may in fact be a CYP2D6 metabolizer-status related effect – the PM vs NM and UM vs NM certainly appear to be quite different. I wonder if the authors have missed an opportunity to conduct a more appropriate statistical test (again, using a more appropriate phenotype) comparing these two outlier groups to NMs, which might in fact reveal an interesting effect of CYP2D6 on oxycodone, which would entirely change the main conclusion of this manuscript.

Our re-analyses using the above major changes has resulted in some differences to the phenotype association effects especially for oxycodone use and RD.

Minor critiques:

5) Hydromorphone is mentioned in the abstract but is not contextualized there, which is therefore confusing.

Hydromorphone has been removed to avoid confusion.

6) Lines 14-18 on page 5 are confusing and self-contradictory and do not accurately reflect current CPIC recommendations.

Reply 6: This discrepancy has been corrected in the manuscript.
Changes in the text: Page 5 lines 12-17.

7) What percent of the genotype-guided group had results available by the time of opioid prescribing?

Reply 7: All patients included in this study had *CYP2D6* genotype data available preoperatively and at the time of opioid prescription.

8) Methods: how were "liver/renal disease" defined? (and to what extend did this exclude patients?)

Reply 8: Patients were identified to have liver and renal disease if they carried such diagnoses in their documented medical history at the time of opioid prescription.
Changes in the text: Page 7 lines 12-13.

9) The standard oxycodone dose on line 1 of page 7 differs from that which is defined as a "normal dose" on page 9 – which is correct?

Patients were identified in clinic and *CYP2D6* testing completed before surgery. Turnaround time is 2 business days and I noted this on page 10 line 4. Page 7 line 8 already states the inclusion criteria: American Society of Anesthesiologists classification 1-3, diagnosis of pectus excavatum or idiopathic scoliosis undergoing Nuss procedure or spine fusion respectively, with *CYP2D6* genotyping results available at the time of surgery.
I have made some comments in text to further clarify it for the reviewers.

10) Page 7: do the authors really mean that an opioid was SCHEDULED every four hours after surgery? That is not consistent with modern analgesia practices, for example, as scheduled dosing leads to risk of opioid-related deaths, and a nurse would need to perform assessments prior to giving the next analgesic. The authors should clarify the protocol being used.

Reply 10: In order to provide more consistent pain control using oral pain medications, nurses were instructed to OFFER opioids to the patients every four hours (scheduled), withholding medications upon patient refusal or patient somnolence. This method of medication prescription circumvents the problems associated with inadequate PRN drug delivery.
Changes in the text: Page 8 line 1.

11) Did opioid prescribing patterns change over the 6 years of this study? The Background says that oxycodone use "recently" became more common – did their data show that? Does this confound their study?

Reply 11: There was no change in opioid prescribing pattern during the study period. Oxycodone had become an integral part of the oral pain medication regimen since 2012, just before this study's data collection timeline.

12) Page 7 Methods say that a clinician receives an alert that a "test" is available – so how long does it take to order the test and then get a result? Does this all happen after the first oral opioid is prescribed? How then would a result return in time for POD2 or 3? What was the genotype turnaround time?

Reply 12: All the patients in this study had *CYP2D6* genotyping performed well before undergoing their respective operation and hence had this data available in the EMR at the time of opioid prescription.
Changes in the text: Page 8 line 2-4.

13) Page 7 mentions tramadol, morphine, and hydrocodone – but nowhere else are these other opioids considered. Were these used in this study? This could potentially introduce a major methodologic confounder.

Reply 13: Mention of tramadol has been removed as it is not part of our practice to prescribe tramadol. Morphine and hydrocodone doses were collected to calculate morphine equivalents and the outcome has been changed to reflect opioid doses used for analgesia (per Reviewer1 comments).

14) RD and Vomiting, as endpoints, are extremely difficult to ascribe as being caused solely by opioids in the post-operative setting. The authors do not acknowledge this. Their data however DO show just how difficult these are to use as endpoints in any formal analysis. The use of these endpoints is highly questioned.

Reply 14: Although RD and vomiting are difficult to solely ascribe to opioid usage, these adverse events are well-known side effects of opioids. We have included several covariates in the univariate analyses to identify other factors that may influence these outcomes, including age, sex, race, surgery type, etc. While none were significant for RD, emesis was higher in patients undergoing pectus surgery and with higher use oxycodone. However, they were not included in multiple regression as they were correlated with phenotypes. Lastly, in order to minimize other confounders leading to RD and vomiting, all patients received scheduled bowel regimens and antiemetics postoperatively.

15) Dichotomization of phenotypes as described on page 9, while perhaps previously done by these authors, is not an accepted method of analysis for CYP2D6 and offers no advantage.

In accordance with the reviewer comments, we have removed the dichotomized phenotypes.

16) Page 10 – "interaction" analysis as described on lines 7-9 (and as later reported in the Results) is not clearly explained and what was actually performed is not clear to this reader.

Thank you for this comment. We have removed interaction analyses from the revised manuscript.

17) The Discussion is too long, and rambles. The authors talk about a lot of prior literature but do not relate the discussion to what they actually found (or did not find).

Reply 17: This has been addressed in the manuscript.

18) Lines 20-22 on page 15 of the Discussion is opinion and editorialization; it is not based in data that they found in this study.

Reply 18: These lines have been removed from the manuscript.

**Reviewer C**

Page 5, last paragraph – clarify that CPIC does not make recommendations about whether to order a pharmacogenetic test, but rather what to do when test results are available. Nonetheless, there is no guidance for oxycodone dosing/administration based on CYP2D6 given the lack of data. I would state that only codeine, tramadol, and to a lesser degree hydrocodone have actionable recommendations.

Reply: This discrepancy has been corrected in the manuscript.

Changes in the text: Page 5 lines 12-17.

Genotyping methods – clarify whether the CN assay can detect allele-specific copy number variations.

Page 10 has been edited to reflect that the assay does not detect allele-specific copy number

Statistical analysis – clarify how AUC was calculated. Why evaluate the interaction effect between CYP2D6 phenotype and hydromorphone when there is no expected interaction (since CYP2D6 does not metabolize hydromorphone)?

Pain experienced over time in the immediate postoperative period is captured well by the area under curve (AUC) which was calculated using trapezoidal rule for area under the curve of pain scores over a time period (as we have used in prior studies).

The authors make several claims of association in instances where the 95% CI is very wide and crosses 1 (e.g. page 12 lines 12 and 22). I would remove these claims of higher odds.

Overall, the paper is lengthy for a largely negative study. The data can be easily illustrated in tables with simple references to the negative results in the table versus typing out each negative result within the text. Additionally, there are many tests/ associations performed and it is easy for the reader to get lost in the text (particularly when there are so many abbreviations).

I don't think it's necessary to repeat p-values in the discussion, especially since everything was negative.

Results have been considerably shortened and results in tables are not repeated in the text.

Was there a practice change based on these negative results (i.e. removal of recommendations to avoid oxycodone in PM/UM patients who undergo clinical genotyping)?

Reply: Based on the findings of this study, we have stopped performing clinical genotyping for *CYP2D6*

The discussion is also quite lengthy and reads more like a review paper than a discussion. Some of the prior data could be better synthesized and put into context of how it compares to this study.

Discussion has been shortened and made more focused.

Should there be a conclusion?

Reply: Conclusion added to the manuscript.

Change in the text: Page 17 line 10-18

**Reviewer D**

Major comments:

- The study population, outcome measures, and statistical comparisons performed are all not adequately described. Are we only looking at pain score changes in patients prescribed oxycodone or also other drugs? What is AUC%? What do the geometric means of AUC% in Table 4 represent?

Thanks for your review. We have further described the population, defined outcome measures and detailed statistical comparisons. We have calculated morphine equivalents and redefined outcome to be relative oxycodone use (compared to total) for analgesia. AUC% is the % of the area under curve of pain scores over a time period corresponding to pain scores. We have removed compositional analysis descriptions and hence geometric means since they are difficult to understand and are not contributing to our study.

- Further, for a study with largely negative findings, it is unclear if it was sufficiently powered for the primary outcome.

This project was designed as a pilot study and hence the sample sizes were comparatively small for evaluation of genetic polymorphisms that are relatively rare such as in PM and UM groups. This has been acknowledged as one of our study's limitations.

- The results are presented in a non-intuitive and hard-to-apply way (Table 4), and non-significant associations are reported as significant findings.

Results have been simplified with a focus on significant findings.

- CPIC guideline recommendations misrepresented in introduction (they do not consider oxycodone clinically actionable)

Reply: This discrepancy has been corrected in the manuscript.

Changes in the text: Page 5 lines 12-17.

Detailed comments:

Abstract:

- Unclear if investigators are examining patients prescribed oxycodone, hydromorphone, or any analgesic after procedure.

All opioid doses given during the period have been considered and morphine equivalents of *CYP2D6* dependent opioid (oxycodone) have been calculated and studied as outcome for analgesia.

- "Adjusted multivariate regression did not find pain area under curve% for mild, moderate and severe pain, respiratory depression or emesis
to be associated with CYP2D6 genetic phenotypes (p=0.565, 0.151, 0.936 respectively) or functional binary phenotypes (p=0.783, 0.408, 0.955 respectively). "

Is this only in patients on oxycodone? Not clear what the three "functional binary phenotypes" refers to.

We have removed the use of functional phenotypes and by creating an outcome of oxycodone/total MEq, we have studied association of CYP2D6 phenotypes with oxycodone relative non oxycodone opioid use.

- Unclear what comparison these p values refer to. Is it PM vs. NM, IM vs. NM, UM vs. NM? Or are the 3 p values for comparisons of patients with mild, moderate, or severe pain at baseline? VERY unclear.

The multiple regression has been reanalyzed and p-values further clarified.

- "higher incidence of side effects and lower AUC for mod-severe pain in PM"... This is post-hoc and not significant, so it is inappropriate to state this bluntly in the abstract.

Thanks. We have removed mention of any non-significant associations or trends.

- Overall lack of clarity. Would recommend reducing number of findings described in the abstract and making sure the methods & results of the abstract are crystal clear about what groups are being compared.

Methods and results have been clarified and simplified.

Intro:

- Strange italics and incorrect citation numbering on Page 4, line 7-9

The italics have been removed and citation addressed.

- Page 5, Line 14 is referring to CPIC guidelines (missing that word)

Reply: “guidelines” added to the line.

Changes in the text: Page 5 line12.

- Page 5, line 17 misrepresents CPIC guideline. Oxycodone does not have the "highest potential for actionability", in fact it is not considered actionable. Codeine and tramadol are the most actionable. This is a major problem and the manuscript is not publishable unless this is corrected.

Reply: This discrepancy has been corrected in the manuscript.

Changes in the text: Page 5 lines 12-17.

Methods:

- Page 9, line 11, unclear if phenoconversion is being considered here. If not, why not?

Reply: In this study we looked for phenoconversion however due to low incidence in our sample size (only two patients in the entire study), no further emphasis was made on its occurrence.

- Page 9, line 20, unclear what AUC% is vs. AUC.

AUC% is the % of pain area under the curve over time, between or below certain pain scores.

- Page 9, line 20, unclear what AUC% for mild, moderate, and severe pain refers to. Aren't pain scores fluctuating?

This has been removed (It refers to %AUC of a curve with time on x-axis and pain

scores on y- axis)

Results:

- page 11, line 16 terribly difficult to read.

This has been removed.

- page 11, 11, line 21. Again, is "AUC% for pain score>4" referring to patients who had pain >4 afer epidural removed but before opioid treatment?

With change in outcome as suggested by reviewer 1, this line has been removed.

- Page 12, line 1-3 very unclear what is being communicated here and whether it is statistically significant.

- Page 12, line 12-13, 95% confidence interval for the OR crosses 1, so no you did not significantly higher difference in risk of RD for PMs.

- Page 12, line 22-23, same issue

- Page 13, line 17, this is not significant. In addition, what is the logistic regression adjusting for?

All the above sentences have been revised extensively.

Table 4: difficult to interpret. What do the geometric means columns mean? Is the ordinal nature of the CYPD26 phenotype in regression model 1 properly reflected in the model? I don't think so, but would recommend statistical review here.

Geometric means associated with compositional analysis have been removed.

Figure 1: percentages for different metabolic pathways don't add up, which confuses reader instead of clarifying. Renal elimination percentages are also a point of confusion.

The percentages provided add up to about 80%. Since bioavailability of oxycodone is 60-70%, we believe the % are reflective of the metabolism reported in the various sources.