OBSTETRICS & GYNECOLOGY

NOTICE: This document contains correspondence generated during peer review and subsequent revisions but before transmittal to production for composition and copyediting:

- Comments from the reviewers and editors (email to author requesting revisions)
- Response from the author (cover letter submitted with revised manuscript)*

**The corresponding author has opted to make this information publicly available.*

Personal or nonessential information may be redacted at the editor's discretion.

Questions about these materials may be directed to the *Obstetrics & Gynecology* editorial office: obgyn@greenjournal.org.

Date: Mar 20, 2020
To: "Brett David Einerson" brett.einerson@hsc.utah.edu
From: "The Green Journal" em@greenjournal.org
Subject: Your Submission ONG-20-476

RE: Manuscript Number ONG-20-476

Cost of elective labor induction versus expectant management in nulliparous women

Dear Dr. Einerson:

Your manuscript has been reviewed by the Editorial Board and by special expert referees. Although it is judged not acceptable for publication in Obstetrics & Gynecology in its present form, we would be willing to give further consideration to a revised version.

If you wish to consider revising your manuscript, you will first need to study carefully the enclosed reports submitted by the referees and editors. Each point raised requires a response, by either revising your manuscript or making a clear and convincing argument as to why no revision is needed. To facilitate our review, we prefer that the cover letter include the comments made by the reviewers and the editor followed by your response. The revised manuscript should indicate the position of all changes made. We suggest that you use the "track changes" feature in your word processing software to do so (rather than strikethrough or underline formatting).

Your paper will be maintained in active status for 21 days from the date of this letter. If we have not heard from you by Apr 10, 2020, we will assume you wish to withdraw the manuscript from further consideration.

REVIEWER COMMENTS:

Reviewer #1: The authors noted in this small subset of patients from the Arrive trial that the total cost of elective labor induction vs expect of management did not differ significantly. It is a well written and the methodology is presented clearly. Interestingly these results challenge the current literature which most show an increase cost from induction. Although they used a complex rubric in their cost methodology that is unfamiliar to me (I have used TreeAge) appears to have been validated in other studies, whether they are similar to the current study is questionable. The analysis appears to be able to drill down on supplies fixed and semi fixed cost of overhead. I am not aware of another study that has been performed at this detailed level.

I have several concerns about the study:

There is inherent bias in your study group. It is a subset (UTAH patients only) of a previously performed trial (ARRIVE) where cost was not one of the primary or secondary outcomes. Some of the more costly events may require larger sample size(neonatal M&M, maternal M&M,etc) In addition, the ARRIVE trial itself has inherent bias as well, the patient's appeared younger, higher rate of minorities, and increase rates of obesity. Interestingly, in the original trial only 27% of eligible women consented to be in the study. Did you find a similar characteristic in your study group compared to all patient's you care for at your hospitals?

You need to provide more data on your patient groups. What percentage of your patients underwent a cesarean section, gestational age of delivery, percentage of expectant management patients that went to induction, and how many patients were delivered after 41 weeks (it was 25% in the ARRIVE trial.) To determine cost point the gestational age upon entry to the study needs to be determned. Costs may be significantly different in the expectant group who were recruited at 38 weeks as opposed to those that are recruited at 38 6/7 weeks?

In your group the only significant difference is in the outpatient antenatal care (OAC) which had a 47% decrease cost in the induction group (unique to your study) and inpatient intrapartum delivery which had an increase of 17% in cost in the same group (which is very similar in previous study). The cost analysis in these 2 areas need to be better developed. As stated above the time of entry into the study, the gestational age at delivery in the expectant group, those that went past 41 weeks in the expectant group would have a significant effect on the cost. I noted from the original trial (ARRIVE) that there was not a protocol for the antepartum/postpartum care in those patients in the expectant group, thus the variability in this management may greatly affect the cost (additional visits, labs, ultrasounds, nonstress test, etc). Please comment?

Did you consider the operational cost? If this was more widespread and accepted by both the population (80% opt in, unlike ARRIVE's 27%) and practicing OB GYNs (not in a small study protocol), would this require additional staff, facility adjustments, etc. Are there indirect costs that one should consider? Please comment?

Although I trust that your cost analysis was done as described, it is very difficult to appreciate the numbers without actual dollar cost. This I feel is a major concern of your study.

Your discussion should include a compare/contrast of a current study in the literature that used a larger study group but used different cost methodology (TreeAge software,PMID:30768934).

In addition
Ln 72: I do not think it is appropriate to list the clinical trial registration since this was not a primary or secondary outcomes of the original NCT 01990612

Ln 95: Please include the above mention study (PMID:30768934)

Ln 147-50: Although you state that the ABC methodology has been used in prior studies, do think those studies are similar in capturing cost as what is seen in Labor and delivery? The previous studies include only pediatrics (sinusitis, a non procedure role diagnosis) and diagnostic testing pediatrics. Are there any surgical examples that has used this cost methodology?

Ln 164: This I feel is very problematic it is difficult to make any assessment without looking at the actual cost?

Ln 173: Since this was an intention to treat, how many in the induction or expectant management group delivered before 39 weeks?

Ln 182: I am confused about the covariant that was used to control for the differences in the ABC methodologies? What is your confidence that this was an appropriate modifier? Did do run any actual patient costs data on individual patients?

Ln 216: Did you re-evaluate your power calculations since you did not have the required 1250 patients (1200)?

Ln 241: Interestingly many costs in the antepartum and postpartum care may be covered by a global fee by many insurance companies, Please comment?

Ln 264: I do not think you have adequately demonstrated this, I think your study suggests that there not statistically significant difference in cost. Although a 7% difference if one considers 1 million deliveries can add up quickly.

Ln 277: Did your analysis taken to affect anesthesia costs?

Ln 281: Would be nice if we had the difference in cost in dollars for your study, as seen in other studies?

Ln 287: In your sub analysis did you notice similar number of cases hypertensive disorders and neonatal support etc? Was your study adequately a powered to find these differences?

Ln 304: If the cost was similar in the different hospitals why was a modifier used when the cost methodology was applied?

Figure 2 is unnecessary.

Reviewer #2: This is a high-quality study on an important question with implications for ob practice--and the US health care system. I applaud the authors on their work.

Abstract
- Could you make the abstract more accessible to a non-technical audience? You do a good job of breaking down the methods in the methods section, but the abstract remains highly complex, such as "advanced time-derived activity-based costing analytics." A shorter methods description might be clearer.
- Does the power calculation need to be in the abstract?

Methods
- The analysis had a solid design with intent-to-treat analysis and the addition of a health system-level variable to account for clustering. Did you consider using a hospital-level clustering variable in addition to the health system clustering variable?
- Thank you for including power calculations to this level of detail.
- You present a good example of how relative costs are calculated, but it was hard to follow how costs were calculated, especially for Intermountain health care. Could you provide an example of how the costing was done for a prenatal visit or similar service to illustrate?
- Would you consider doing a sensitivity analysis going to 60 days postpartum to coincide with current Medicaid coverage period?

Results
- Thank you for including the Utah vs. non-Utah results from the ARRIVE trial for context.
- Did you analyze cost differences for women who had IOL > vaginal delivery vs. IOL > C-section? Understanding the cost difference with mode of delivery might be helpful, especially in places that find they have a higher C-section rate with IOL than seen in the ARRIVE trial.
- Similarly, it would be interesting to know the cost differences for women with severe maternal morbidity and/or perinatal death. It might help clinicians (and payers) better counsel patients and allocate resources.
- Would you consider presenting more data than the effect modification p-value for your effect modification analysis? Maybe an Appendix Table showing your generalized linear regression model.

Discussion
- The authors present a good synthesis of the literature and strengths/limitations of their study.

Reviewer #3: The authors conducted a cost analysis of a subset of the ARRIVE trial. I have the following suggestions for improvement:

1. Abstract: Instead of percent differences, could the authors provide the actual estimated costs for the different groups? $X vs. $Y?

2. Introduction:

3. Methods: The costing mechanisms of the two health systems are different, did the authors compare the results in the two systems to ascertain whether there were differences?

4. Methods: The bit about actual costs not available for publication is unreasonable in my opinion. You are not reporting individual's costs, but those for a large group. If the data is important to have published, then we should know the results.

5. Methods: Why were there missing cost data? Given that it is health system data, wouldn't they have costs for all patients?

6. Methods: For missing data, why not do imputation?

7. Methods: Statistical analysis seems very thoughtful. Appreciate the use of GLM model. Not sure I entirely understand the use of the Breslow-Day test given that you didn't have cost data for other sites.

8. Methods: The a priori power calculation assume normalcy of the data? I ask because you mention standard deviations. But, the data was non-normal and was not analyzed as if normal, so was that the appropriate power calculation?

9. Methods: Given that the number of patients available (just over 1200) were already known, why did the authors conduct a power calculation for 1250?

10. Methods: Is there anyway for the authors to look at extremes of costs? That is, were there differences in rates above some high threshold, e.g. $30,000?

11. Results: Tough not to have the actual numbers.

12. Results: The 5% increase has a p-value of 0.18 and seems underpowered.

13. Results: Another item that they might report/analyze are the estimated costs to charge ratio - that would be an important bit of information and would allow future investigators to estimate costs.

14. Discussion: While on an individual basis, costs were not statistically different between enrolled patients, one question that remains is whether the additional 3-4 hours on labor and delivery would stress the system leading to either worse outcomes in other patients or higher costs in other patients? For example, if L&D is full, would patients be delayed leading to higher costs?

15. Discussion: Any comment on the power issue post hoc given comments above and that there was a 5% increase in total costs numerically?

STATISTICAL EDITOR COMMENTS:

The Statistical Editor makes the following points that need to be addressed:
lines 163-165: While it is understandable that reporting of individual costs for tests or services might be limited, aggregation of such data to to compare mean costs would not seem to violate any individual's privacy nor any proprietary information. Particularly if costs are being compared, a difference of 4.7% needs context to be interpreted properly or being compared to the sample size calculation offered.

lines 215-218: Should provide a separate Table showing the mean (and SD) for overall costs and for the various subsets of costs listed in Table 2. It seems that the overall cost was the primary outcome of interest, so that should be separated clearly in Table 2 from the subsets, which are all secondary outcomes. Also, if the previous calculations are applicable to this cohort, then it appears that the study was under powered (ie, < 80% power) to discern a difference of 4.7%. A better context for the results would be to list the mean costs (primary, then secondary outcomes) along with their 95% CIs for the two groups. That would allow the reader to see the mean cost and plausible range of costs for the two groups.

Regarding the adjusted mean difference of 4.7% higher costs for IOL, if the same relative difference was replicable in a larger cohort, then the CI would decrease and the difference, although modest, would not be nothing. The population level cost differential could be substantial, if replicated on a larger scale.

lines 231-238: What were the actual counts or rates of these adverse outcomes? Were the counts and total sample sizes sufficiently powered to generalize the NS findings?

lines 256-259: Were there sufficient sample sizes, ie, adequate power, to test for these interaction terms?

lines 240-243: Should report the unadjusted as well as the adjusted mean differences. It is unclear why adjustment was required, since the groups were randomized and the Authors demonstrated no difference by health system.

EDITOR'S COMMENTS:

We no longer require that authors adhere to the Green Journal format with the first submission of their papers. However, any revisions must do so. I strongly encourage you to read the instructions for authors (the general bits as well as those specific to the feature-type you are submitting). The instructions provide guidance regarding formatting, word and reference limits, authorship issues, and other things. Adherence to these requirements with your revision will avoid delays during the revision process, as well as avoid re-revisions on your part in order to comply with the formatting. For instance, in your methods section you have subheadings which we do not use.

44: The précis is a single sentence of no more than 25 words, written in the present tense and stating the conclusion(s) of the report (ie, the bottom line). The précis should be similar to the abstract conclusion. Do not use commercial names, abbreviations, or acronyms in the précis. Precis should be the "hook" for people who scan the Table of Contents to see what to read. It shouldn't not include statements like "in this study" or "we found". Just state what you found.

49: Would you consider “To compare …..at 39 weeks to cost of expectant management”.

51: Your methods section isn’t completely clear in that the ARRIVE trial was multicenter and you are just looking at five sites’ economic results, which needs to be clear. Also, sentence starting w/ “Economic analysis” is not a complete sentence.

79: Would you consider adding “non-obstetrically indicated (elective induction) at 39 weeks….” As you know, a lot of the confusion around induction of labor, whether following the ARRIVE trial or related to what has become known at the “39 week Rule”, revolves around distinguishing between elective and indicated inductions. Just to add clarity and as a reminder, I think the suggested terminology (or something similar of your choosing) would be helpful.

96: If you are going to make sweeping comments like “used suboptimal methodologies to determine actual costs” you should provide some references to support them.

118: They were randomized from 38 week 0 days to 38 weeks 6 days not between them (which would exclude the two anchor ages). You wrote the induction times on line 120 correctly.

129: you don’t need to include IRB/consent sentence for the ARRIVE trial, but the IRB approval for this economic analysis needs to be kept.

162: Since the methods for accounting for costs at the two systems, Intermountain and Univ. of Utah are different, how did you combine them for this study?

163: All of your reviewers and I agree that your paper would be much stronger if the actual aggregated dollar amounts were provided. It should be possible to get the total costs in dollars in the two arms for each phase of care and report that, along w/ the percent differences. You could aggregate them for the 5 sites, so if the costs at the two sites is different, then that would not be exposed.

205: Was the outpatient care for included patients (prenatal, post partum and infant care) all provided in settings at which the Health System accounting processes worked? At some hospitals, independent private practices or even practices that are part of health care systems, may not be a part of the health care system's accounting process but admit patients for delivery at the hospital. Please clarify that this was NOT the case in your participants or if it was the case, how did you include data from these patients?

216: Power calculation: I am a little confused by the way you used a power calculation. You had a fixed number of potential participants (1201); thus, you cannot enroll enough patients for this power calculation and your results are therefore a priori underpowered. Why did you not do a post hoc power calculation with the fixed number of patients to see what percent difference, with an 85% power, you could detect? That would actually tell the reader one measure of potential validity of your analysis.

245: As a reminder to the reader, could you say "The cost of outpatient antenatal care from the time of enrollment was 47% lower…."

266: As you don't provide the data that the results were true in both participating health systems you should probably say that in parentheses here.

Starting w/ line 278: Please clarify if these studies reported on total OB and Neonatal care (In patient and outpatient) or only some segment of that.

284: The Green Journal uses the reVITALize definitions, adopted by ACOG. Please use these definitions. For instance, the correct terminology for PROM is prelabor rupture of the membranes.

Philosophically, I think it is important to add a comment in your discussion about how a patient, ob practice, or hospital/health care system might consider your results. Your analysis was from the perspective of the health care system. As my question above notes, for some women, antenatal and postnatal care will be provided in settings that are financially separate from the birthing hospital. The patient who is either paying out of pocket or has a deductible/copays to deal with, might look at your data be motivated to decrease outpatient expenses. Her ob practice might be motivated to embrace elective induction to decrease the costs of her care (under global reimbursement) while the hospital might be motivated to limit this practice given the higher intrapartum costs. The different "buckets" and lenses that entities use to make these decisions warrants some comment, in my opinion. This is not a deal breaker for your paper, but I think including some comment about this would be valued by the readers.

EDITORIAL OFFICE COMMENTS:

1. The Editors of Obstetrics & Gynecology are seeking to increase transparency around its peer-review process, in line with efforts to do so in international biomedical peer review publishing. If your article is accepted, we will be posting this revision letter as supplemental digital content to the published article online. Additionally, unless you choose to opt out, we will also be including your point-by-point response to the revision letter. If you opt out of including your response, only the revision letter will be posted. Please reply to this letter with one of two responses:

A. OPT-IN: Yes, please publish my point-by-point response letter.
B. OPT-OUT: No, please do not publish my point-by-point response letter.

2. As of December 17, 2018, Obstetrics & Gynecology has implemented an "electronic Copyright Transfer Agreement" (eCTA) and will no longer be collecting author agreement forms. When you are ready to revise your manuscript, you will be prompted in Editorial Manager (EM) to click on "Revise Submission." Doing so will launch the resubmission process, and you will be walked through the various questions that comprise the eCTA. Each of your coauthors will receive an email from the system requesting that they review and electronically sign the eCTA.

Please check with your coauthors to confirm that the disclosures listed in their eCTA forms are correctly disclosed on the manuscript's title page.

3. Responsible reporting of research studies, which includes a complete, transparent, accurate and timely account of what was done and what was found during a research study, is an integral part of good research and publication practice and not an optional extra. Obstetrics & Gynecology supports initiatives aimed at improving the reporting of health research, and we ask authors to follow specific guidelines for reporting randomized controlled trials (ie, CONSORT), observational studies (ie, STROBE), meta-analyses and systematic reviews of randomized controlled trials (ie, PRISMA), harms in systematic reviews (ie, PRISMA for harms), studies of diagnostic accuracy (ie, STARD), meta-analyses and systematic reviews of observational studies (ie, MOOSE), economic evaluations of health interventions (ie, CHEERS), quality

improvement in health care studies (ie, SQUIRE 2.0), and studies reporting results of Internet e-surveys (CHERRIES). Include the appropriate checklist for your manuscript type upon submission. Please write or insert the page numbers where each item appears in the margin of the checklist. Further information and links to the checklists are available at http://ong.editorialmanager.com. In your cover letter, be sure to indicate that you have followed the CONSORT, MOOSE, PRISMA, PRISMA for harms, STARD, STROBE, CHEERS, SQUIRE 2.0, or CHERRIES guidelines, as appropriate.

4. Standard obstetric and gynecology data definitions have been developed through the reVITALize initiative, which was convened by the American College of Obstetricians and Gynecologists and the members of the Women's Health Registry Alliance. Obstetrics & Gynecology has adopted the use of the reVITALize definitions. Please access the obstetric and gynecology data definitions at https://www.acog.org/About-ACOG/ACOG-Departments/Patient-Safety-and-Quality-Improvement/reVITALize. If use of the reVITALize definitions is problematic, please discuss this in your point-by-point response to this letter.

5. Because of space limitations, it is important that your revised manuscript adhere to the following length restrictions by manuscript type: Original Research reports should not exceed 22 typed, double-spaced pages (5,500 words). Stated page limits include all numbered pages in a manuscript (i.e., title page, précis, abstract, text, references, tables, boxes, figure legends, and print appendixes) but exclude references.

6. Specific rules govern the use of acknowledgments in the journal. Please note the following guidelines:

* All financial support of the study must be acknowledged.
* Any and all manuscript preparation assistance, including but not limited to topic development, data collection, analysis, writing, or editorial assistance, must be disclosed in the acknowledgments. Such acknowledgments must identify the entities that provided and paid for this assistance, whether directly or indirectly.
* All persons who contributed to the work reported in the manuscript, but not sufficiently to be authors, must be acknowledged. Written permission must be obtained from all individuals named in the acknowledgments, as readers may infer their endorsement of the data and conclusions. Please note that your response in the journal's electronic author form verifies that permission has been obtained from all named persons.
* If all or part of the paper was presented at the Annual Clinical and Scientific Meeting of the American College of Obstetricians and Gynecologists or at any other organizational meeting, that presentation should be noted (include the exact dates and location of the meeting).

7. The most common deficiency in revised manuscripts involves the abstract. Be sure there are no inconsistencies between the Abstract and the manuscript, and that the Abstract has a clear conclusion statement based on the results found in the paper. Make sure that the abstract does not contain information that does not appear in the body text. If you submit a revision, please check the abstract carefully.

In addition, the abstract length should follow journal guidelines. The word limits for different article types are as follows: Original Research articles, 300 words. Please provide a word count.

8. Only standard abbreviations and acronyms are allowed. A selected list is available online at http://edmgr.ovid.com/ong/accounts/abbreviations.pdf. Abbreviations and acronyms cannot be used in the title or précis. Abbreviations and acronyms must be spelled out the first time they are used in the abstract and again in the body of the manuscript.

9. The journal does not use the virgule symbol (/) in sentences with words. Please rephrase your text to avoid using "and/or," or similar constructions throughout the text. You may retain this symbol if you are using it to express data or a measurement.

10. In your Abstract, manuscript Results sections, and tables, the preferred citation should be in terms of an effect size, such as odds ratio or relative risk or the mean difference of a variable between two groups, expressed with appropriate confidence intervals. When such syntax is used, the P value has only secondary importance and often can be omitted or noted as footnotes in a Table format. Putting the results in the form of an effect size makes the result of the statistical test more clinically relevant and gives better context than citing P values alone.

If appropriate, please include number needed to treat for benefits (NNTb) or harm (NNTh). When comparing two procedures, please express the outcome of the comparison in U.S. dollar amounts.

Please standardize the presentation of your data throughout the manuscript submission. For P values, do not exceed three decimal places (for example, "P = .001"). For percentages, do not exceed one decimal place (for example, 11.1%").

11. Please review the journal's Table Checklist to make sure that your tables conform to journal style. The Table Checklist is available online here: http://edmgr.ovid.com/ong/accounts/table_checklist.pdf.

12. The American College of Obstetricians and Gynecologists' (ACOG) documents are frequently updated. These documents may be withdrawn and replaced with newer, revised versions. If you cite ACOG documents in your manuscript, be sure the reference you are citing is still current and available. If the reference you are citing has been updated (ie, replaced by a newer version), please ensure that the new version supports whatever statement you are making in your manuscript and then update your reference list accordingly (exceptions could include manuscripts that address items of

historical interest). If the reference you are citing has been withdrawn with no clear replacement, please contact the editorial office for assistance (obgyn@greenjournal.org). In most cases, if an ACOG document has been withdrawn, it should not be referenced in your manuscript (exceptions could include manuscripts that address items of historical interest). All ACOG documents (eg, Committee Opinions and Practice Bulletins) may be found via the Clinical Guidance & Publications page at https://www.acog.org/Clinical-Guidance-and-Publications/Search-Clinical-Guidance.

13. Figures 1 and 2 may be resubmitted with the revision.

14. Authors whose manuscripts have been accepted for publication have the option to pay an article processing charge and publish open access. With this choice, articles are made freely available online immediately upon publication. An information sheet is available at http://links.lww.com/LWW-ES/A48. The cost for publishing an article as open access can be found at http://edmgr.ovid.com/acd/accounts/ifauth.htm.

Please note that if your article is accepted, you will receive an email from the editorial office asking you to choose a publication route (traditional or open access). Please keep an eye out for that future email and be sure to respond to it promptly.

15. If you choose to revise your manuscript, please submit your revision through Editorial Manager at http://ong.editorialmanager.com. Your manuscript should be uploaded in a word processing format such as Microsoft Word. Your revision's cover letter should include the following:

* A confirmation that you have read the Instructions for Authors (http://edmgr.ovid.com/ong/accounts/authors.pdf), and
* A point-by-point response to each of the received comments in this letter.

If you submit a revision, we will assume that it has been developed in consultation with your co-authors and that each author has given approval to the final form of the revision.

Again, your paper will be maintained in active status for 21 days from the date of this letter. If we have not heard from you by Apr 10, 2020, we will assume you wish to withdraw the manuscript from further consideration.

Sincerely,

Nancy C. Chescheir, MD
Editor-in-Chief

2018 IMPACT FACTOR: 4.965
2018 IMPACT FACTOR RANKING: 7th out of 83 ob/gyn journals

_____

To the Editors of *Obstetrics & Gynecology*,

Thank you for this opportunity to revise our manuscript entitled "Cost of elective labor induction versus expectant management in nulliparous women." We appreciate the work and perspective of the reviewers and editors who responded to our initial submission.

Please find below our responses to reviewer and editor remarks.

With regard to the peer review process. OPT-IN: Yes, please publish my point-by-point response letter.

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained. IRB 00100258.

I have read the Instructions for Authors.

Thank you for your consideration.

Sincerely,

Brett D. Einerson, MD MPH

University of Utah Health [redacted]

REVIEWER COMMENTS:

Reviewer #1:

The authors noted in this small subset of patients from the Arrive trial that the total cost of elective labor induction vs expect of management did not differ significantly. It is a well written and the methodology is presented clearly. Interestingly these results challenge the current literature which most show an increase cost from induction. Although they used a complex rubric in their cost methodology that is unfamiliar to me (I have used TreeAge) appears to have been validated in other studies, whether they are similar to the current study is questionable. The analysis appears to be able to drill down on supplies fixed and semi fixed cost of overhead. I am not aware of another study that has been performed at this detailed level.

RESPONSE: Thank you for these comments.

I have several concerns about the study:

There is inherent bias in your study group. It is a subset (UTAH patients only) of a previously performed trial (ARRIVE) where cost was not one of the primary or secondary outcomes. Some of the more costly events may require larger sample size(neonatal M&M, maternal M&M,etc) In addition, the ARRIVE trial itself has inherent bias as well, the patient's appeared younger, higher rate of minorities, and increase rates of obesity. Interestingly, in the original trial only 27% of eligible women consented to be in the study. Did you find a similar characteristic in your study group compared to all patient's you care for at your hospitals?

RESPONSE: We acknowledge the above critiques of the original ARRIVE trial and this Utah cohort. Some of these have been well documented elsewhere, in editorials and letters responding to the trial. Likewise, we address the limitations of our cohort in the discussion.

We did not perform an analysis to evaluate this cohort versus all patients in our hospital. We contend that there is no reason to believe that minor differences between low risk nulliparous women who enrolled and low risk nulliparous women who did not enroll would significantly alter the relative cost between IOL and EM.

We echo a recent editorial by Dr. Collins and colleagues that "in generalizing the results of a randomized trial, the assumption is not that the patient population studied is representative of all patients but rather that *the proportional effects* of the treatment studied on each specific health outcome *should be similar* in different circumstances *unless there is a good reason to expect otherwise*" (*emphasis added*) (Collins. NEJM 2020;382:674-678). Here, there is not that expectation.

You need to provide more data on your patient groups. What percentage of your patients underwent a cesarean section, gestational age of delivery, percentage of expectant management patients that went to induction, and how many patients were delivered after 41 weeks (it was 25% in the ARRIVE trial.) To determine cost point the gestational age upon entry to the study needs to be determned. Costs may be significantly different in the expectant group who were recruited at 38 weeks as opposed to those that are recruited at 38 6/7 weeks?

RESPONSE: The MFMU policy does not allow for "per site" data. Through tests of interaction, we have demonstrated that the clinical outcomes of Utah are similar to those of the other ARRIVE sites.

In your group the only significant difference is in the outpatient antenatal care (OAC) which had a 47% decrease cost in the induction group (unique to your study) and inpatient intrapartum delivery which had an increase of 17% in cost in the same group (which is very similar in previous study). The cost analysis in these 2 areas need to be better developed. As stated above the time of entry into the study, the gestational age at delivery in the expectant group, those that went past 41 weeks in the expectant group would have a significant effect on the cost. I noted from the original trial (ARRIVE) that there was not a protocol for the antepartum/postpartum care in those patients in the expectant group, thus the variability in this management may greatly affect the cost (additional visits, labs, ultrasounds, nonstress test, etc). Please comment?

RESPONSE: We too are interested in determining whether labor management strategies or other factors contribute to cost escalation in this cohort. This was not the primary focus of this study, and doing so would be well beyond the scope of this analysis, which is focused on the primary question of whether one strategy (IOL) is different in cost than another (EM). Nevertheless, we plan to explore this in more detail in future analyses.

Did you consider the operational cost? If this was more widespread and accepted by both the population (80% opt in, unlike ARRIVE's 27%) and practicing OB GYNs (not in a small study protocol), would this require additional staff, facility adjustments, etc. Are there indirect costs that one should consider? Please comment?

RESPONSE: Facility and overhead costs are included in the activity-based costing (ABC) methodologies used in this study (Kawamoto 2015), but cannot be completely accounted on a population level. The indirect economic effects of implementing a policy of elective labor (outside of the trial) are unknown. As we state in the discussion: "we cannot know whether these costs will be similar among women undergoing care outside of a trial setting, and in particular with volume or patient flow differences that hospitals may experience after implementing a policy allowing elective induction."

Although I trust that your cost analysis was done as described, it is very difficult to appreciate the numbers without actual dollar cost. This I feel is a major concern of your study.

RESPONSE: Please see our first response to the Statistical Editor and Editor below.

Your discussion should include a compare/contrast of a current study in the literature that used a larger study group but used different cost methodology (TreeAge software,PMID:30768934).

RESPONSE: We added a new paragraph in the discussion section to highlight differences between our study and the Hersh study. Please note that "larger study group" is not a relevant critique since the Hersh study is a cost-effectiveness modeling exercise. In contrast, we performed measurement of actual patient costs within a randomized trial.

Line 318 – 340: "Our study stands in contrast to a recent cost-effectiveness analysis by Hersh and colleagues concluding that elective induction is likely to add considerable cost ($2 billion annually) to the U.S. health system.[12] Differing conclusions can be

explained by several differences in study design. First, cost-effectiveness modeling is not the same thing as cost measurement. We measured and compared the actual costs of health care for each Utah patient enrolled in the trial. Cost-effectiveness modeling as performed in the Hersh et al analysis does not measure actual costs accumulated throughout the study directly, but uses simulation modeling to compare the cost and clinical effectiveness of different approaches for patient care. These simulation models are parameterized using estimates derived from the published literature or other data sources. As a result, Hersh and co-authors included costs associated with induction (+$1,979.87 (range $1000-$3000) that were higher than actual costs we measured. Second, Hersh modeled costs of rare events not found to be different in the ARRIVE trial (e.g. neonatal death and brachial plexus injury) over an entire life span for both the newborn and mother, whereas our study was limited to the costs produced during the study period from 38 weeks to 6-8 weeks postpartum. Third, Hersh evaluated cost from a societal standpoint, which is appropriate for cost-effectiveness modeling, but would not be appropriate or possible for our study's main objective of measuring actual costs in a health system."

In addition
Ln 72: I do not think it is appropriate to list the clinical trial registration since this was not a primary or secondary outcomes of the original NCT 01990612

RESPONSE: This has been removed.

Ln 95: Please include the above mention study (PMID:30768934)

RESPONSE: This is now included as reference 12.

Ln 147-50: Although you state that the ABC methodology has been used in prior studies, do think those studies are similar in capturing cost as what is seen in Labor and delivery? The previous studies include only pediatrics (sinusitis, a non procedure role diagnosis) and diagnostic testing pediatrics. Are there any surgical examples that has used this cost methodology?

RESPONSE: The tools are used by both institutions on services throughout the hospitals and clinics involved. It is widely used, validated, and improved internally for value-improvement projects, and has yielded multiple publications at each institution. A surgical example from the Intermountain system is now included.

Reference 16: Meier JD, Duval M, Wilkes J, Andrews S, Korgenski EK, Park AH, Srivastava R. Surgeon dependent variation in adenotonsillectomy costs in children. Otolaryngol Head Neck Surg 2014;150(5):887-892.

Ln 164: This I feel is very problematic it is difficult to make any assessment without looking at the actual cost?

RESPONSE: We want to be sure that the reader understands: our analysis is a comparison of actual cost. While reporting the absolute values of these costs would be informative, we are unable to do so – and can instead only present the comparison of actual costs. Please see responses to the editors below. As a current policy, neither institution was willing to share or publish these actual cost data.

The following clarification has been made:

Line 189 and following: "It is the policy of both Intermountain and UUH to limit the granularity of cost data available for publication and dissemination. As such, the actual costs of individual tests and services or the absolute values of overall costs cannot be reported (though they were measured and compared)."

Additionally, we discuss the fact that *actual costs* were compared throughout the manuscript:

Line 118: "Thus, the objective of this study was to *measure and compare the actual health care costs* of induction and expectant management in women enrolled in the ARRIVE trial at Utah sites where detailed cost data were measured."

Line 192: "Instead of absolute overall costs, we report the relative difference *in actual measured costs* between arms in the trial: elective labor induction and expectant management… For example, a cost difference of +10% signifies actual costs that are 10% higher with induction, whereas a cost difference of -10% signifies actual costs that are 10% lower with induction."

Line 231: "We measured and compared the actual costs of health care for each patient enrolled in the trial…"

Line 376: "…our *costs were measured among actual women* participating in a randomized trial, reducing the potential for bias inherent in observational and modeling studies.

Line 392: "Nevertheless, the relative costs based on *actual measured dollar amounts* still provide important comparative insight for health systems…

Ln 173: Since this was an intention to treat, how many in the induction or expectant management group delivered before 39 weeks?

RESPONSE: The MFMU policy is that outcomes by site are not published separate from the primary trial. We therefore did not report this outcome.

Ln 182: I am confused about the covariant that was used to control for the differences in the ABC methodologies? What is your confidence that this was an appropriate modifier? Did do run any actual patient costs data on individual patients?

RESPONSE: The covariate is used to control for differences in the mean costs between institutions. Regarding the last question, yes, all of the costs included here are actual patient costs on individual patients. If the editors would like us to include more language on this point, please let us know where it would be most effective and clear.

Ln 216: Did you re-evaluate your power calculations since you did not have the required 1250 patients (1200)?

RESPONSE: Please see the revised post hoc power calculation.

Line 294: <u>"A post-hoc power analysis demonstrated that with 1,201 patients with a two-sided alpha of 0.05 we had a power of 85% to detect a 7.3% difference in mean cost between randomization arms"</u>

Ln 241: Interestingly many costs in the antepartum and postpartum care may be covered by a global fee by many insurance companies, Please comment?

RESPONSE: Our analysis was from the perspective of the health system, but other perspectives matter, particularly when different parties bear the burden of cost. Please see our added commentary in response to the editor's last comment:

Line 403: "<u>it is important to acknowledge that this analysis is from just one economic perspective – the health system. Patients facing out-of-pocket copays may prefer an approach that minimizes additional visits and testing. Hospitals with critical staffing shortages may prioritize a strategy that limits patient time on the labor ward. Clinicians who receive reimbursement for outpatient procedures and visits may see the reduced antenatal costs with induction as possible revenue lost."</u>

Ln 264: I do not think you have adequately demonstrated this, I think your study suggests that there not statistically significant difference in cost. Although a 7% difference if one considers 1 million deliveries can add up quickly.

RESPONSE: We agree that a 7% cost difference (in favor of either arm) is considerable on a population level. Our study had power to detect differences in cost much smaller than previously thought to exist between induction and expectant management. Strictly speaking we found no statistical difference, and the confidence interval included cost savings in favor of induction *and* cost escalation in favor of expectant management. We have clarified this with the following language in the limitations section in response to the statistical editor's comment:

Line 396: "<u>although we found no statistical difference in costs in this study, real cost differences below the threshold our study was powered to detect (+/- 7.3%) may still be relevant. If, in contrast to our findings, real cost differences exist, the population cost or savings of elective labor induction could be substantial."</u>

Ln 277: Did your analysis taken to affect anesthesia costs?

RESPONSE: Yes. This has been clarified in the methods.

Line 235: "inpatient intrapartum and delivery care (<u>including anesthesia</u>) from the time of admission for delivery through time of vaginal or cesarean delivery"

Ln 281: Would be nice if we had the difference in cost in dollars for your study, as seen in other studies?

RESPONSE: Please see our response to the Statistical Editor and Editor below regarding reporting of actual costs.

Ln 287: In your sub analysis did you notice similar number of cases hypertensive disorders and neonatal support etc? Was your study adequately a powered to find these differences?

RESPONSE: The original comment, referenced here, was one about the parent ARRIVE trial. The MFMU does not allow for single site outcome data to be published separately. We evaluated differences in outcomes and the interaction term for those comparisons is presented. As an economic analysis, we did not evaluate whether we had adequate power to find differences in these outcomes just in Utah participants.

Ln 304: If the cost was similar in the different hospitals why was a modifier used when the cost methodology was applied?

RESPONSE: Costs were not the same between institutions, but the effect of the randomization arm on costs was similar. Clarification was added to the discussion.

Line 374: "the results were obtained from several different hospitals (including academic and community centers) that care for patients with different demographic characteristics, have different payor mixes, have different staff providing obstetric care, had different mean costs (not reported), and have different obstetric volumes. These consistent results lend support to the generalizability of these findings."

Additionally, the line following that sentence has been removed for the sake of clarity: ~~"Not only were the overall combined costs similar, they were similar within each health care system as well."~~

Figure 2 is unnecessary.

RESPONE: We think this is an important figure since these data are not presented elsewhere. The figure gives readers a visual representation of the absolute differences in costs per phase (and between arms), which is not well-represented elsewhere. The figure can be omitted, at the editors' request.

Reviewer #2:

This is a high-quality study on an important question with implications for ob practice--and the US health care system. I applaud the authors on their work.

RESPONSE: Thank you for these comments.

Abstract
- Could you make the abstract more accessible to a non-technical audience? You do a good job of breaking down the methods in the methods section, but the abstract remains highly complex, such as "advanced time-derived activity-based costing analytics." A shorter methods description might be clearer.

RESPONSE: The phrase "time-derived activity-based" has been removed.

- Does the power calculation need to be in the abstract?

RESPONSE: In our opinion, a "negative" study should be contextualized with a statement of power and sample size adequacy. This seems in line with the comments made by the statistical editor, but we can remove this if needed.

Methods
- The analysis had a solid design with intent-to-treat analysis and the addition of a health system-level variable to account for clustering. Did you consider using a hospital-level clustering variable in addition to the health system clustering variable?

RESPONSE: We considered this, but had planned an analysis that accounted for differences in health system as this would be the major source of cost variation.

- Thank you for including power calculations to this level of detail.

- You present a good example of how relative costs are calculated, but it was hard to follow how costs were calculated, especially for Intermountain health care. Could you provide an example of how the costing was done for a prenatal visit or similar service to illustrate?

RESPONSE: The time-derived activity-based costing (ABC) methods used at each institution are very similar in their approach.

The following has been added for clarification:

Line 175: "Data are harvested from the health system's enterprise data warehouse, incorporating all relevant input from patient encounters <u>in a process similar to methods used by Intermountain</u>. The tool calculates the cost of an encounter by first tallying the number of units of supplies, pharmacy, imaging, and laboratory components of care and then applying a unit cost by cross-referencing to the institution's complete record of financial transactions…"

- Would you consider doing a sensitivity analysis going to 60 days postpartum to coincide with current Medicaid coverage period?

RESPONSE: We chose the time frame (out to 6-8 weeks) based on the parent ARRIVE trial. We expect that few patients would interact with the health system after 6-8 weeks (42-56 days) and before the 60 day mark.

Results
- Thank you for including the Utah vs. non-Utah results from the ARRIVE trial for context.

- Did you analyze cost differences for women who had IOL > vaginal delivery vs. IOL > C-section? Understanding the cost difference with mode of delivery might be helpful, especially in places that find they have a higher C-section rate with IOL than seen in the ARRIVE trial.

RESPONSE: See response below.

- Similarly, it would be interesting to know the cost differences for women with severe maternal morbidity and/or perinatal death. It might help clinicians (and payers) better counsel patients and allocate resources.

RESPONSE: We too are interested in drivers of cost - particularly those that may be modifiable. Costs attributable to specific services, outcomes, or practices were not evaluated in this analysis, as that is an enormous and complex analysis well beyond the scope of this manuscript. Here, we focused upon the important question of whether there are differences in cost between the two strategies examined. We hope to report in the future on drivers of cost in this cohort.

- Would you consider presenting more data than the effect modification p-value for your effect modification analysis? Maybe an Appendix Table showing your generalized linear regression model.

RESPONSE: We believe that presentation of interactions terms (lines 225-230, and line 290-293) here is adequate to convey the point that each of these pre-defined clinical variables did not modify the effect of randomization on cost. We did not originally provide a table of the GLMs, as these may result in mis-interpretation of non-significant trends in these subgroups that would be a distraction from the primary objective here.

Discussion
- The authors present a good synthesis of the literature and strengths/limitations of their study.

Reviewer #3:

The authors conducted a cost analysis of a subset of the ARRIVE trial. I have the following suggestions for improvement:

1. Abstract: Instead of percent differences, could the authors provide the actual estimated costs for the different groups? $X vs. $Y?

RESPONSE: See our response to the statistical editor and editor below.

2. Introduction:

3. Methods: The costing mechanisms of the two health systems are different, did the authors compare the results in the two systems to ascertain whether there were differences?

RESPONSE: The aim of the study was to evaluate difference between arms, not institutions. Our data sharing agreement requests that differences in cost between these two competing health systems not be reported.

4. Methods: The bit about actual costs not available for publication is unreasonable in my opinion. You are not reporting individual's costs, but those for a large group. If the data is important to have published, then we should know the results.

RESPONSE: Please see our response to the statistical editor and editor below regarding actual cost reporting. Mean costs of an entire group can be used by payors in negotiating bundled payments. Because the negotiated costs of inputs into patient care

between suppliers and hospitals are confidential, we are prohibited from revealing the actual dollar amounts associated with the encounters described in our paper.

The methods used here are well established and considered to be the gold standard for cost measurement in health economics (Kawamoto et al. J Am Med Inform Assoc 2015;22:223-235 and Barnett. Med Care 2009;47:S82-88), and reporting of normalized or relative costs is used elsewhere (e.g. Lee et al, JAMA 2016:316:1061-1072).

5. Methods: Why were there missing cost data? Given that it is health system data, wouldn't they have costs for all patients?

RESPONSE: Some patients did not receive all of their antenatal and postpartum care within the reach of the accounting systems involved (e.g. had prenatal care or follow-up outside of the two health systems). They were excluded.

A clarification has been added to the methods:

Line 140. "Participants with missing or incomplete cost data (if their prenatal, postpartum or infant care were outside of the two institutions) were excluded."

6. Methods: For missing data, why not do imputation?

RESPONSE: This could have been done, though was deemed unnecessary given how few patients (2%) had missing cost data.

7. Methods: Statistical analysis seems very thoughtful. Appreciate the use of GLM model. Not sure I entirely understand the use of the Breslow-Day test given that you didn't have cost data for other sites.

RESPONSE: The reviewer may be referring to the Breslow-Day tests performed to evaluate whether the Utah site participants differed in their outcomes from the main trial. This was not a comparison of cost differences.

8. Methods: The a priori power calculation assume normalcy of the data? I ask because you mention standard deviations. But, the data was non-normal and was not analyzed as if normal, so was that the appropriate power calculation?

RESPONSE: A revised post hoc power calculation has been reported in the revision. See line 294.

9. Methods: Given that the number of patients available (just over 1200) were already known, why did the authors conduct a power calculation for 1250?

RESPONSE: This was the pre-analysis power calculation based on projected enrollment (before trial enrollment finished). This has been removed. A revised post hoc power calculation has been reported in the revision.

10. Methods: Is there anyway for the authors to look at extremes of costs? That is, were there differences in rates above some high threshold, e.g. $30,000?

RESPONSE: There were no differences between IOL and EM in percentage of patients with extreme costs >$30,000, (0.5% vs 0.5%, p = 0.99) or in costs >$20,000 (1.9% vs 2.2%, p=0.71). This can be added to the results as a post hoc analysis, if requested.

11. Results: Tough not to have the actual numbers.

RESPONSE: Please see our response to the statistical editor and editor below regarding actual cost reporting.

12. Results: The 5% increase has a p-value of 0.18 and seems underpowered.

RESPONSE: Our discussion now includes a more-thorough discussion of power.

Line 394: "although we found no statistical difference in costs in this study, real cost differences below the threshold our study was powered to detect (+/- 7.3%) may still be relevant. If, in contrast to our findings, real cost differences exist, the population cost or savings of elective labor induction could be substantial."

We add also:

Line 400: "Until this time, elective labor induction has been thought to be clearly more expensive than EM (by 17-50%).[6-8,12] Our analysis of actual cost suggests this difference is not likely."

13. Results: Another item that they might report/analyze are the estimated costs to charge ratio - that would be an important bit of information and would allow future investigators to estimate costs.

RESPONSE: As cost-to-charge ratios differ by region, by institution, by department, and even by test, it is not clear to us how this would clarify our results. Relative costs can easily be applied to future investigations to estimate absolute costs. In addition, because charges do not reflect the cost of providing care to patients nor do they reflect the cost of that care to payers we feel they are not terribly useful to present here.

14. Discussion: While on an individual basis, costs were not statistically different between enrolled patients, one question that remains is whether the additional 3-4 hours on labor and delivery would stress the system leading to either worse outcomes in other patients or higher costs in other patients? For example, if L&D is full, would patients be delayed leading to higher costs?

RESPONSE: This is an important question that our study cannot fully address. Our discussion addressed this point.

Line 383: "we cannot know whether these costs will be similar among women undergoing care outside of a trial setting, and in particular with volume or patient flow differences that hospitals may experience after implementing a policy allowing elective induction."

We add to the discussion:

Line 403: "<u>it is important to acknowledge that this analysis is from just one economic perspective – the health system.</u> … <u>Hospitals with critical staffing shortages may prioritize a strategy that limits patient time on the labor ward</u>."

15. Discussion: Any comment on the power issue post hoc given comments above and that there was a 5% increase in total costs numerically?

RESPONSE: Please see our response to reviewer 3, question 12.

STATISTICAL EDITOR COMMENTS:

The Statistical Editor makes the following points that need to be addressed:
lines 163-165: While it is understandable that reporting of individual costs for tests or services might be limited, aggregation of such data to to compare mean costs would not seem to violate any individual's privacy nor any proprietary information. Particularly if costs are being compared, a difference of 4.7% needs context to be interpreted properly or being compared to the sample size calculation offered.

RESPONSE: We agree that mean or median costs would be informative. It is not our intention to be opaque, and there is no scientific reason to conceal these actual cost data. I, the first author, can personally assure the editors and readers that we have gone to all lengths, over more than a year of negotiation with the legal teams from these two competing health systems, to make these costs as transparent as possible. Both systems have an interest to conceal actual cost to some degree, as payers (insurers) could use published data to renegotiate bundled payments for pregnancy care. If we were to report the mean cost of labor and postpartum care, we would essentially reveal the great majority of the cost of an entire pregnancy episode, which, in our region, is negotiated between payers and providers as a bundled payment. This level of cost transparency is not generally allowed by either institution's current policies, and was not specifically allowed for this study. Additionally, fully revealing and comparing costs could identify cost differences between two health systems competing in the same market.

I would urge one other point of clarity: cost for the same service (e.g. cesarean) is known to vary widely from health system to health system. With regard to induction vs expectant management, a non-significant cost difference of +4.7% (95% CI -2.1% to +12.0%) can easily be applied to health systems whether their cost of this episode in a given health system is $1,500, or $15,000. We offer that the generalizable result from this study is not the absolute difference in costs in our specific health systems, but the relative cost difference that may be applied to other health systems with varying absolute costs (e.g. Lee et al JAMA 2016;316:1061-1072).

lines 215-218: Should provide a separate Table showing the mean (and SD) for overall costs and for the various subsets of costs listed in Table 2. It seems that the overall cost was the primary outcome of interest, so that should be separated clearly in Table 2 from the subsets, which are all secondary outcomes. Also, if the previous calculations are applicable to this

cohort, then it appears that the study was under powered (ie, < 80% power) to discern a difference of 4.7%. A better context for the results would be to list the mean costs (primary, then secondary outcomes) along with their 95% CIs for the two groups. That would allow the reader to see the mean cost and plausible range of costs for the two groups.

RESPONSE: Overall costs have been removed from Table 2, as they are mentioned in the text. For the above noted reasons, we cannot report cost means, medians, or standard deviations. If there is another way to satisfy this request, please let us know.

Regarding the adjusted mean difference of 4.7% higher costs for IOL, if the same relative difference was replicable in a larger cohort, then the CI would decrease and the difference, although modest, would not be nothing. The population level cost differential could be substantial, if replicated on a larger scale.

RESPONSE: This point is well taken. We have added the following to the limitations.

Line 395. "Fourth, although we found no statistical difference in actual costs in this study, cost differences below the threshold our study was powered to detect (+/- 7.3%) may still be relevant. If, in contrast to our findings, real cost differences exist, the population cost or savings of elective labor induction could be substantial."

lines 231-238: What were the actual counts or rates of these adverse outcomes? Were the counts and total sample sizes sufficiently powered to generalize the NS findings?

RESPONSE. See response below.

lines 256-259: Were there sufficient sample sizes, ie, adequate power, to test for these interaction terms?

RESPONSE The outcome for this analysis was actual cost, not event rates. Tests of interaction were therefore performed on event rates to give the reader some reassurance that major differences in outcomes do not exist between the Utah cohort and the other MFMU sites.

lines 240-243: Should report the unadjusted as well as the adjusted mean differences. It is unclear why adjustment was required, since the groups were randomized and the Authors demonstrated no difference by health system.

RESPONSE: Adjustment is necessary because the mean costs are different between systems. We demonstrated no difference in the relative cost between arms *by health system*.

As stated in the results (emphasis added): "The mean total *cost of induction compared to expectant management did not differ* by health system (interaction p=0.74)."

Line 273: We have now reported the unadjusted relative cost differences: "The mean total cost of induction was not significantly different than that of expectant management, (adjusted mean difference +4.7%, 95% CI -2.1% to +12.0%, p=0.18; unadjusted mean difference +5.0%, 95% CI -2.1% to +12.5%)."

EDITOR'S COMMENTS:

We no longer require that authors adhere to the Green Journal format with the first submission of their papers. However, any revisions must do so. I strongly encourage you to read the instructions for authors (the general bits as well as those specific to the feature-type you are submitting). The instructions provide guidance regarding formatting, word and reference limits, authorship issues, and other things. Adherence to these requirements with your revision will avoid delays during the revision process, as well as avoid re-revisions on your part in order to comply with the formatting. For instance, in your methods section you have subheadings which we do not use.

RESPONSE: We have removed all subheadings, and have reformatted this revision to Green Journal standards.

44: The précis is a single sentence of no more than 25 words, written in the present tense and stating the conclusion(s) of the report (ie, the bottom line). The précis should be similar to the abstract conclusion. Do not use commercial names, abbreviations, or acronyms in the précis. Precis should be the "hook" for people who scan the Table of Contents to see what to read. It shouldn't not include statements like "in this study" or "we found". Just state what you found.

RESPONSE: We have rewritten the precis.

Line 44: <u>"Total actual costs of elective labor induction do not differ significantly from expectant management among term, nulliparous women enrolled in a randomized controlled trial."</u>

49: Would you consider "To compare …..at 39 weeks to cost of expectant management".

RESPONSE: This has been revised.

Line 53: <u>"To compare the actual health-system cost of elective labor induction at 39 weeks to expectant management."</u>

51: Your methods section isn't completely clear in that the ARRIVE trial was multicenter and you are just looking at five sites' economic results, which needs to be clear. Also, sentence starting w/ "Economic analysis" is not a complete sentence.

RESPONSE: This has been clarified.

Line 55: "<u>This was an</u> economic analysis <u>of patients enrolled in the five Utah hospitals participating in a multicenter</u> randomized trial of elective labor induction at 39 weeks versus expectant management in low-risk nulliparous women. <u>The entire trial enrolled over 6,000 patients. For this subset, 1,201 had cost data available</u>."

79: Would you consider adding "non-obstetrically indicated (elective induction) at 39 weeks…." As you know, a lot of the confusion around induction of labor, whether following the ARRIVE trial or related to what has become known at the "39 week Rule", revolves around

distinguishing between elective and indicated inductions. Just to add clarity and as a reminder, I think the suggested terminology (or something similar of your choosing) would be helpful.

RESPONSE: Thank you. This has been clarified.

Line 91: "A Randomized Trial of Induction Versus Expectant Management (ARRIVE) was a landmark multicenter trial of over 6,000 low-risk nulliparous women comparing non-obstetrically indicated induction of labor (elective induction) at 39 weeks gestation with expectant management."

96: If you are going to make sweeping comments like "used suboptimal methodologies to determine actual costs" you should provide some references to support them.

RESPONSE: Thank you. Three references have been added that describe the limitations of other costing methodologies and specifically identify the methodology we used (ABC or time-derived ABC) as the gold standard or "the best available estimate of economic costs of health care services."

Ref 9. Barnett PG. An improved set of standards for finding cost for cost-effectiveness analysis. Med Care 2009;47:S82-88.
Ref 10. Kaplan RS, Witkowski M, Abbott M, et al. Using time-driven activity-based costing to identify value improvement opportunities in healthcare. J Healthc Manag. 2014;59(6):399-412.
Ref 11. Kaplan RS, Porter ME. How to solve the cost crisis in health care. Harv Bus Rev 2011;89(9):46-61.

118: They were randomized from 38 week 0 days to 38 weeks 6 days not between them (which would exclude the two anchor ages). You wrote the induction times on line 120 correctly.

RESPONSE: This has been corrected.
Line 130. "Participants were randomized from 38 weeks 0 days gestation to 38 weeks 6 days gestation to induction of labor or expectant management."

129: you don't need to include IRB/consent sentence for the ARRIVE trial, but the IRB approval for this economic analysis needs to be kept.

RESPONSE: Clinical trials registration number has been removed.

162: Since the methods for accounting for costs at the two systems, Intermountain and Univ. of Utah are different, how did you combine them for this study?

RESPONSE: The methods and means differ slightly, but both methods measure the exact same outcome – actual direct health care costs. Differences in means between the institutions were controlled for in generalized linear modeling.

We clarified in the methods:

Line 162: "Costs were measured using advanced analytics platforms at each health system. Although each health system has slightly different costing methodology, they both used activity-based costing (ABC) to assign direct episode-based costs to each patient individually. In this regard, they both measure the same outcome."

…and in the discussion:

Line 374: "Second, the results were obtained from several different hospitals (including academic and community centers) that care for patients with different demographic characteristics, have different payor mixes, have different staff providing obstetric care, had different mean costs (not reported), and have different obstetric volumes."

163: All of your reviewers and I agree that your paper would be much stronger if the actual aggregated dollar amounts were provided. It should be possible to get the total costs in dollars in the two arms for each phase of care and report that, along w/ the percent differences. You could aggregate them for the 5 sites, so if the costs at the two sites is different, then that would not be exposed.

RESPONSE: We agree that this is a limitation and that presenting actual cost data would be ideal. We simply cannot report these data per each institution's current policies and our data-sharing agreement for this study. See response to the first comment by the Statistical Editor.

205: Was the outpatient care for included patients (prenatal, post partum and infant care) all provided in settings at which the Health System accounting processes worked? At some hospitals, independent private practices or even practices that are part of health care systems, may not be a part of the health care system's accounting process but admit patients for delivery at the hospital. Please clarify that this was NOT the case in your participants or if it was the case, how did you include data from these patients?

RESPONSE: This is an important point. In our hospitals, most patients enrolled in the trial at outpatient clinics that are operated by the two hospital systems. Only patients with complete cost data from all phases were included in this economic evaluation. Thus, for the great majority (1,201 of 1,231), complete inpatient and outpatient data were available for maternal and neonatal care.

A clarification has been added to the methods:

Line 140. "Participants with missing or incomplete cost data (if their prenatal, postpartum or infant care were outside of the two institutions) were excluded."

216: Power calculation: I am a little confused by the way you used a power calculation. You had a fixed number of potential participants (1201); thus, you cannot enroll enough patients for this power calculation and your results are therefore a priori underpowered. Why did you not do a post hoc power calculation with the fixed number of patients to see what percent difference, with an 85% power, you could detect? That would actually tell the reader one measure of potential validity of your analysis.

RESPONSE: We originally reported the power calculation used for the grant proposal (before full enrollment of the trial had completed, based on projected Utah enrollment statistics). To be more informative, we have recalculated a post-hoc power calculation with the fixed numbers in our dataset, and placed it in the results (not methods):

Line 294: "A post-hoc power analysis demonstrated that with 1,201 patients and a two-sided alpha of 0.05 we had a power of 85% to detect at least 7.3% difference in mean cost between randomization arms."

245: As a reminder to the reader, could you say "The cost of outpatient antenatal care from the time of enrollment was 47% lower…."

RESPONSE: This has been changed.

Line 280: "When considering each phase of care, the cost of outpatient antenatal care from the time of enrollment to labor admission was 47% lower…"

266: As you don't provide the data that the results were true in both participating health systems you should probably say that in parentheses here.

RESPONSE: This statement has been removed.

Starting w/ line 278: Please clarify if these studies reported on total OB and Neonatal care (In patient and outpatient) or only some segment of that.

RESPONSE: These studies did not consistently include all components of care and varied in their horizons and perspective (see below).

Ref 22 (now 28). Vligen (2010) – Costs from societal standpoints. Maternal and neonatal. 1 year time horizon (inpatient and outpatient) costs up to the time of discharge. Economic analysis of HYPITAT.

Ref 23 (now 29). Walker – Costs from NHS (health system) perspective. Maternal only. Antenatal, intrapartum, and postpartum(inpatient and outpatient) costs from randomization to 4 weeks postpartum. Economic analysis of AMA IOL Walker Trial.

Ref 24 (now 30) – Vligen (2013) – Costs from health system perspective. Maternal and neonatal. Antenatal, intrapartum, and postpartum costs (inpatient and outpatient) up to the time of discharge. Economic analysis of DIGITAT trial.

Ref 25 (now 31) – Vligen (2014) – Costs from health system perspective. Maternal and neonatal. Antenatal, intrapartum, and postpartum (inpatient and outpatient) costs up to the time of discharge. Economic analysis of PROMEXIL.

The following has been added to that paragraph:

Line 348: "Some of these studies did not include outpatient costs or neonatal costs, or used a cost perspective other than the health system perspective."

284: The Green Journal uses the reVITALize definitions, adopted by ACOG. Please use these definitions. For instance, the correct terminology for PROM is prelabor rupture of the membranes.

RESPONSE: We have changed premature to prelabor.

Philosophically, I think it is important to add a comment in your discussion about how a patient, ob practice, or hospital/health care system might consider your results. Your analysis was from the perspective of the health care system. As my question above notes, for some women, antenatal and postnatal care will be provided in settings that are financially separate from the birthing hospital. The patient who is either paying out of pocket or has a deductible/copays to deal with, might look at your data be motivated to decrease outpatient expenses. Her ob practice might be motivated to embrace elective induction to decrease the costs of her care (under global reimbursement) while the hospital might be motivated to limit this practice given the higher intrapartum costs. The different "buckets" and lenses that entities use to make these decisions warrants some comment, in my opinion. This is not a deal breaker for your paper, but I think including some comment about this
would be valued by the readers.

RESPONSE: Thank you for this comment, and for the opportunity to expand our discussion on this topic. Before the concluding paragraph of the discussion we have added the following.

Line 403: "Finally, from a philosophical standpoint, it is important to acknowledge that this analysis is from just one economic perspective – the health system. Patients facing out-of-pocket copays may prefer an approach that minimizes additional visits and testing. Hospitals with critical staffing shortages may prioritize a strategy that limits patient time on the labor ward."