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: 12/23/2022
To: "Adi Hirshberg" [redacted]
From: "The Green Journal" em@greenjournal.org
Subject: Your Submission ONG-22-1756

RE: Manuscript Number ONG-22-1756

A remote blood pressure monitoring program's impact on postpartum adverse outcomes

Dear Dr. Hirshberg:

Thank you for sending us your work for consideration for publication in Obstetrics & Gynecology. Your manuscript has been reviewed by the Editorial Board and by special expert referees. The Editors would like to invite you to submit a revised version for further consideration.

If you wish to revise your manuscript, please read the following comments submitted by the reviewers and Editors. Each point raised requires a response, by either revising your manuscript or making a clear argument as to why no revision is needed in the cover letter.

To facilitate our review, we prefer that the cover letter you submit with your revised manuscript include each reviewer and Editor comment below, followed by your response. That is, a point-by-point response is required to each of the EDITOR COMMENTS (if applicable), REVIEWER COMMENTS, and STATISTICAL EDITOR COMMENTS (if applicable) below.

The revised manuscript should indicate the position of all changes made. Please use the "track changes" feature in your document (do not use strikethrough or underline formatting).

Your submission will be maintained in active status for 21 days from the date of this letter. If we have not heard from you by 01/13/2023, we will assume you wish to withdraw the manuscript from further consideration.

EDITOR COMMENTS:

Please note the following:

* Help us reduce the number of queries we add to your manuscript after it is revised by reading the Revision Checklist at https://journals.lww.com/greenjournal/Documents/RevisionChecklist_Authors.pdf and making the applicable edits to your manuscript.

* Please make sure you are responsive to the comments from that Statistical Editor in your revision.

REVIEWER COMMENTS:

Reviewer #1: I enjoyed reading your manuscript. I have a few comments that I hope may serve to strengthen it.

In LNs 27 - 30 & LNs 148 - 151, you discuss lower cost i.e. cost per member per month. Are you certain this is a result of the intervention, or could other factors have influenced the finding? In the monitoring to asynchronous comparison, it would be important to know if the reimbursement rates for the same services were identical over the multiyear period as reimbursement rates frequently fluctuate over time. In the monitoring to contemporaneous comparison, it would be important to know if the hospitals utilizing the monitoring program were reimbursed at identical rates to those that did not utilize the program as different hospitals or health systems are frequently reimbursed at different rates by the same insurance company.

While your analysis was performed on administrative data, I wondered if you have any information regarding program implementation costs. In your discussion you note that this and societal costs require future evaluation, but while reading the manuscript I kept wondering how the implementation costs would effect the reduction in cost per member per month.

In LNs 59 - 63, "However, this study was a cost-analysis based on inputs and assumptions extrapolated from literature,

focused on readmission rates. The purpose of this study was to evaluate the impact of a text-message based remote blood pressure monitoring program on adverse postpartum clinical outcomes and costs in patients with a hypertensive disorder of pregnancy." The word "this" being used twice was a bit confusing. I believe that with the first "this" you are referring to the prior study you are referencing, and with the second "this" the current study you are presenting, but the diction is a bit confusing, and you may wish to consider clarifying it.

In Tables 2 and 3, You report data showing that the patients in the remote monitoring program experienced a statistically significant reduction in total postpartum complications during the first six months postpartum in comparison to the patients in each matched group. In your manuscript, you did not mention, but I found it interesting, that the patients in the remote monitoring program experienced fewer complications than the patients in both the asynchronous and contemporaneous groups in each of the 8 categories of postpartum complications that you evaluated. I wondered if you might consider pointing that out.

Reviewer #2: In this study, the authors sought to determine if remote home BP monitoring in the first 10 days postpartum was associated with a reduction of adverse clinical outcomes in the year following delivery, in women with hypertensive disorders of pregnancy.

The authors compared a cohort of women with hypertensive disorders of pregnancy who performed home BP monitoring with two separate control cohorts: a historical (asynchronous) group who delivered in the 2 years prior to the subject cohorts and a concurrent cohort of women with hypertensive disorders of pregnancy who delivered at different hospitals than the subject cohort. Neither of these control cohorts performed home BP monitoring.

Adverse outcomes included CVA, DIC, eclampsia, pulmonary edema, HELLP syndrome, myocardial infarction, and cardiomyopathy. These adverse outcomes, as well as information regarding hospitalizations and health care utilization in the year after delivery, were not evaluated by review of clinical data, but rather by an analysis of administrative claims (health insurance claims (Independence Blue Cross) and ICD-10 codes).

The authors enrolled 1700 members in the remote home BP monitoring program.

Comparisons were made with 1021 asynchronous controls and with 1276 concurrent controls.

The authors found that there were significantly more postnatal complications in both of the control groups. There was only one complication that occurred in any of the controls and subjects (one of the asynchronous controls) more than 6 months after delivery.

Additionally, the authors found that subjects had significantly less ER visits and hospital readmissions than patients in either of the control groups. However, subjects were noted to have significantly more specialty and cardiology visits in the year following delivery.

The authors conclude that a 10 day post delivery program of home BP evaluation is associated with a decrease in hypertension-related adverse outcomes in the year after delivery.

Critique/questions:
In most centers, postpartum evaluation of blood pressure has become a routine component of post delivery care in women with hypertensive disorders of pregnancy.

This study attempts to illustrate that home BP assessment reduces adverse outcomes in the first year after delivery and is cost saving.

1. Can the authors provide some description of how patients were instructed to measure blood pressure. Was it multiple times/day? Specific times? If a patient was delivered by CS and went home POD 4 - did she perform BP evaluations for a full10 days? What were the target BP thresholds that were utilized in this study to signify the need for intervention?

2. There is very limited clinical information on the women in any of the cohorts and in the women who experienced an adverse outcome. Clinical information that would be important includes: BMI, gravidity and parity, use of low dose aspirin, gestational age at delivery, route of delivery, antihypertensive therapy (what agents, how many agents), use of other medications, need for peripartum magnesium sulfate, laboratory abnormalities prior to and after delivery. This information would be valuable to gauge the severity of the hypertensive disorder and the comparability of the subjects and controls. Is it conceivable that the adverse outcomes occurred only/primarily in women with more severe hypertensive disorders and those that required preterm delivery? Is it possible that women with more mild disease (e.g., chronic hypertension that required no treatment during pregnancy, gestational hypertension that did not progress to preeclampsia) don't benefit from home BP evaluation? Is it possible that women who experienced serious cardiovascular disorders (CVA, MI, cardiomyopathy) were at greater risk for a postpartum complication?

3. The authors performed a "risk" score when matching the cohorts and indicate that there were no differences in cardiovascular conditions (appendices 1 and 2). What were these cardiovascular conditions?

4. Can the authors provide more clear information on the timing of the adverse events in the subjects and controls? Although these complications occurred after delivery, did all of these happen after discharge or did some of these complications occur after delivery but in the hospital prior to discharge. Some of the complications that were analyzed are ones that are most commonly associated with a complication at or soon after delivery - DIC, pulmonary edema, renal injury/liver failure, for example. If some of these complications occurred postnatally but prior to hospital discharge, it is unclear how home BP assessment after discharge is relevant. Furthermore, if all of these complications occurred within the first 1-2 months postpartum, it may be inappropriate to conclude that home BP evaluation is associated with reduced complications for a year after delivery.

5. The term HELLP syndrome is often used a bit too liberally and often erroneously to describe a patient with preeclampsia with mild laboratory abnormalities. The most common postpartum hypertensive complication that requires readmission is postpartum preeclampsia and this was not one of the adverse outcomes that was evaluated. Are the authors using the term HELLP syndrome synonymously with the term preeclampsia? If not, why wasn't readmission for preeclampsia studied?

6. Although there were less ER visits and hospital readmissions in the control groups, there is no information on the reason(s) for these visits. If these were visits for a gynecologic problem (e.g., miscarriage, STI) or a URI/pneumonia, for example, how would a postpartum home BP evaluation be relevant to these issues?

7. This study was conducted in the subjects and concurrent cohorts from 9.2017-4.2021. There is an established correlation between Covid19 infection and preeclampsia [AJOG Sept 2021;e1-17]. Do the authors have any information on the incidence of Covid19 in their subjects and controls? Did the incidence of hypertensive disease increase in the final year of enrollment (table 1 shows higher incidence of hypertension in subjects and concurrent controls)? Were adverse postdelivery outcomes more common in women who had Covid during pregnancy? Do the authors feel that patients with Covid19 infection during pregnancy should be candidates for home BP evaluation?

Reviewer #3:

GENERAL COMMENTS: The authors present a retrospective cohort study assessing the impact of a remote blood pressure monitoring program on adverse postpartum outcomes and medical costs. Patients enrolled in twice daily text-based blood pressure monitoring for ten days postpartum (n=1,700) were compared to two propensity score matched cohorts of i) patients at any of the three participating hospitals before remote monitoring program implementation (n=1,021) and ii) patients at other hospitals during the same time period as clinical use of the program (n=1,276). This is an important area of inquiry because postpartum hypertension is an important cause of morbidity and is pernicious because peak blood pressures occur 3 - 6 days postpartum when most patients are already home. Preliminary data from an RCT showed increased blood pressure ascertainment overall and reduced disparities with remote blood pressure monitoring, but the study was not powered to detect differences in outcomes. The results of this study showing that remote blood pressure monitoring reduced postpartum adverse outcomes, readmissions, ER visits and medical costs, and increased cardiology visits up to a year after delivery are very promising. A number of revisions listed below will further strengthen the manuscript.

SPECIFIC COMMENTS:
1. Inclusion criteria (patients with gestational hypertension, pre-eclampsia (PEC), chronic hypertension with superimposed PEC, HELLP syndrome, or eclampsia at the time of their delivery admission) are appropriate, but the exclusion of members who incurred over $12,500 in total medical costs per member per month (PMPM) in the prenatal period is not justified. These are likely the patients with comorbidities and therefore at the highest risks of adverse outcomes.

2. The use of two propensity score-matched control cohorts is reasonable to account for confounders. An additional sensitivity analysis would be to perform a difference-in-difference analysis: i.e. compare change in outcomes before to after implementation of the remote blood pressure monitoring to the change in outcomes from the same before to after implementation timeframe in the hospitals where the program was never implemented. This would account for changes in practice over time.

3. The procedures for the propensity matching appear appropriate.

4. Outcomes were a composite of adverse outcomes identified through ICD-10 codes and healthcare resource utilization. Because not all diagnoses make it to claims documents there could have been under reporting of the adverse outcomes. This should be admitted as a limitation

5. None of the outcomes as designated the primary outcome. Although the issue of statistical power is moot because of

the significant differences seen, it would be important to be explicit on how the investigators arrived at the sample size used.

6. Effect sizes, rather than only p-values, should be presented for the outcomes. And even after propensity score matching, adjusted analysis may still be performed as needed.

7. As disparities are an important feature of maternal morbidity and mortality data in the U.S., analysis should be performed by race/ethnicity to assess if remote blood pressure monitoring is equitable by reducing disparities in adverse outcomes and healthcare utilization.

8. It must be acknowledged that while propensity score matching accounted for measured confounders there is still the possibility of residual confounding from unmeasured variables.

STATISTICAL EDITOR COMMENTS:

lines 29-33: First, the data provided relates to 6 months following delivery, not 1st year. See later comments re: tables 3, 4, 5, since many of the comparisons are NS. The third sentence is not correct, in that this was not a randomized study and therefore the differences cannot be generalized as necessarily applying to future cohort.

lines 80-81: Was enrollment voluntary and if so, how many were enrolled and how many declined or did not complete the "exposure"? Should include a flow diagram.

lines 89-90: The emulation of randomization only applied to known confounders. Since the groups were not randomly allocated, unknown confounders remain unknown.

lines 94-96: Need to provide a reference or supplemental material demonstrating the risk score calculation.

Methods: Need to explain more fully the treatment group BP monitoring, its compliance, frequency, loss to follow-up etc.

Should include a Table similar to Table 1 but comparing the matched cohorts.

Table 2: For this comparison, the rates (2.9% vs 4.7%) are significantly different, but only at a p = 0.04 level. Should include the CIs for each proportion, which are 2.0-4.2% and 3.5-6.2%, respectively.

Table 3: For the contemporaneous cohorts, the difference in rates is NS. They were 3.2% (2.3-4.4%) vs 4.5% (3.4-5.8%), with p = 0.11. Applying other test statistics, Chi-square = 2.74 or OR = 1.41 (0.94-2.12), both of course with same p = 0.10

Table 4: Should change the format to show the actual counts in each matched group as n (%), rather than as rates per 1000. Should include CIs for the differences and omit the column of p-values. For these cohorts, none of the numerical differences are statistically significant. For those with relatively low rates (ER visits or readmits), the results are both NS and vastly underpowered to discern such small differences.

Table 5: Again, should change the format, include CIs and omit the p-values. The results for these contemporaneous groups are each NS.

Should include the covariate balance figures to main text. Suggest omitting the K-S stats comparisons and the remaining text explaining

--
Sincerely,
Jason D. Wright, MD
Editor-in-Chief

The Editors of Obstetrics & Gynecology

Adi Hirshberg, MD

[redacted]

REVIEWER COMMENTS:

Reviewer #1: I enjoyed reading your manuscript. I have a few comments that I hope may serve to strengthen it.

In LNs 27 - 30 & LNs 148 - 151, you discuss lower cost i.e. cost per member per month. Are you certain this is a result of the intervention, or could other factors have influenced the finding? In the monitoring to asynchronous comparison, it would be important to know if the reimbursement rates for the same services were identical over the multiyear period as reimbursement rates frequently fluctuate over time. In the monitoring to contemporaneous comparison, it would be important to know if the hospitals utilizing the monitoring program were reimbursed at identical rates to those that did not utilize the program as different hospitals or health systems are frequently reimbursed at different rates by the same insurance company.

Thank you for highlighting the potential difference in reimbursement rates that could impact the reported outcomes. Regarding year-over-year unit cost increases that are relevant for comparisons to the asynchronous group, we do not adjust for inflation in these analyses. In this instance inflation would bias our finding toward the null. Inflation-adjusted savings would be greater than those reported here. You are correct to point out that contracting rates differ across providers. We use a version of the cost variable that accounts for rate differences across providers.

While your analysis was performed on administrative data, I wondered if you have any information regarding program implementation costs. In your discussion you note that this and societal costs require future evaluation, but while reading the manuscript I kept wondering how the implementation costs would effect the reduction in cost per member per month.

Thank you for this comment. The program implementation costs include the cost of the blood pressure monitor and the platform through which the automated texting occurs. Blood pressure monitors, in general, cost $20-30 per patient; however, these are often obtained during prenatal care and are often covered by insurance, and therefore this cost is often reduced or negligible. The program cost is about $8000 annually across our health system, with about 4900 patients enrolled a year, resulting in cost of about $1.60 per patient when excluding blood pressure monitor costs. Management is done as part of daily clinical responsibilities, resulting in no additional administrative costs. We have not shared these implementation costs publicly but for purposes of this review believe this minimal cost per patient would not significantly reduce the reduction in cost per member per month.

In LNs 59 - 63, "However, this study was a cost-analysis based on inputs and assumptions

extrapolated from literature, focused on readmission rates. The purpose of this study was to evaluate the impact of a text-message based remote blood pressure monitoring program on adverse postpartum clinical outcomes and costs in patients with a hypertensive disorder of pregnancy." The word "this" being used twice was a bit confusing. I believe that with the first "this" you are referring to the prior study you are referencing, and with the second "this" the current study you are presenting, but the diction is a bit confusing, and you may wish to consider clarifying it.

Thank you for pointing out the confusing language. The text has been modified to read "the aforementioned study…" and "The purpose of the current study was to evaluate the impact of a text-message based remote blood pressure monitoring program…." to provide more clarity. This has been changed in lines 57 and 59 respectively.

In Tables 2 and 3, You report data showing that the patients in the remote monitoring program experienced a statistically significant reduction in total postpartum complications during the first six months postpartum in comparison to the patients in each matched group. In your manuscript, you did not mention, but I found it interesting, that the patients in the remote monitoring program experienced fewer complications than the patients in both the asynchronous and contemporaneous groups in each of the 8 categories of postpartum complications that you evaluated. I wondered if you might consider pointing that out.

Thank you for this suggestion. This is an interesting finding that we have highlighted in the text in lines 168-171, which now reads "There were fewer postnatal adverse outcomes among patients enrolled in the remote blood pressure monitoring program across each of the eight categories measured in the six months following delivery, though individually they were not statistically significant (See Tables 2 and 3)." We have also emphasized this finding in the discussion in lines 210-213, pointing out that the decrease in every adverse outcome included complications that are not directly mediated by severe hypertension.

Reviewer #2: In this study, the authors sought to determine if remote home BP monitoring in the first 10 days postpartum was associated with a reduction of adverse clinical outcomes in the year following delivery, in women with hypertensive disorders of pregnancy.

The authors compared a cohort of women with hypertensive disorders of pregnancy who performed home BP monitoring with two separate control cohorts: a historical (asynchronous) group who delivered in the 2 years prior to the subject cohorts and a concurrent cohort of women with hypertensive disorders of pregnancy who delivered at different hospitals than the subject cohort. Neither of these control cohorts performed home BP monitoring.

Adverse outcomes included CVA, DIC, eclampsia, pulmonary edema, HELLP syndrome, myocardial infarction, and cardiomyopathy. These adverse outcomes, as well as information regarding hospitalizations and health care utilization in the year after delivery, were not evaluated by review of clinical data, but rather by an analysis of administrative claims (health insurance claims (Independence Blue Cross) and ICD-10 codes).

The authors enrolled 1700 members in the remote home BP monitoring program.

Comparisons were made with 1021 asynchronous controls and with 1276 concurrent controls.

The authors found that there were significantly more postnatal complications in both of the control groups. There was only one complication that occurred in any of the controls and subjects (one of the asynchronous controls) more than 6 months after delivery.

Additionally, the authors found that subjects had significantly less ER visits and hospital readmissions than patients in either of the control groups. However, subjects were noted to have significantly more specialty and cardiology visits in the year following delivery.

The authors conclude that a 10 day post delivery program of home BP evaluation is associated with a decrease in hypertension-related adverse outcomes in the year after delivery.

Critique/questions:
In most centers, postpartum evaluation of blood pressure has become a routine component of post delivery care in women with hypertensive disorders of pregnancy.

This study attempts to illustrate that home BP assessment reduces adverse outcomes in the first year after delivery and is cost saving.

1. Can the authors provide some description of how patients were instructed to measure blood pressure. Was it multiple times/day? Specific times? If a patient was delivered by CS and went home POD 4 - did she perform BP evaluations for a full10 days? What were the target BP thresholds that were utilized in this study to signify the need for intervention?
We appreciate this comment. Though these are important aspects of the intervention we intentionally omitted these details in the initial submission because the protocol is described in prior published work, which we reference in the current manuscript. However, for the purposes of this paper and value to the reader, we have added a paragraph with details of the remote monitoring program in the methods. These details are expanded upon in lines 78-96.

2. There is very limited clinical information on the women in any of the cohorts and in the women who experienced an adverse outcome. Clinical information that would be important includes: BMI, gravidity and parity, use of low dose aspirin, gestational age at delivery, route of delivery, antihypertensive therapy (what agents, how many agents), use of other medications, need for peripartum magnesium sulfate, laboratory abnormalities prior to and after delivery. This information would be valuable to gauge the severity of the hypertensive disorder and the comparability of the subjects and controls. Is it conceivable that the adverse outcomes occurred only/primarily in women with more severe hypertensive disorders and those that required preterm delivery? Is it possible that women with more mild disease (e.g., chronic hypertension that required no treatment during pregnancy, gestational hypertension that did not progress to preeclampsia) don't benefit from home BP evaluation? Is it possible that women who experienced serious

cardiovascular disorders (CVA, MI, cardiomyopathy) were at greater risk for a postpartum complication?
We thank the reviewer for these observations. The clinical characteristics you mention are indeed important aspects of the severity of the hypertensive disorder. Although the researchers have access to clinical data on all the dimensions the reviewer lists, these data are not reliably available in medical claims data for the comparison cohorts. Our approach was to use the propensity score matching process to control for relevant clinical characteristics for which we had reliable data on both the treatment and comparison cohorts. Although we have no reason to believe a priori that there were differences in the intervention and comparison cohorts on the distribution of hypertensive severity, we cannot conclude that there were no differences. We have noted this limitation in the discussion in lines 253-254.

3. The authors performed a "risk" score when matching the cohorts and indicate that there were no differences in cardiovascular conditions (appendices 1 and 2). What were these cardiovascular conditions?
Thank you for this comment. The cardiac conditions we included were acute myocardial infarction, atrial fibrillation, heart failure, ischemic heart disease, and stroke/transient ischemic attack. We have added those to the Table 1 notes.

4. Can the authors provide more clear information on the timing of the adverse events in the subjects and controls? Although these complications occurred after delivery, did all of these happen after discharge or did some of these complications occur after delivery but in the hospital prior to discharge. Some of the complications that were analyzed are ones that are most commonly associated with a complication at or soon after delivery - DIC, pulmonary edema, renal injury/liver failure, for example. If some of these complications occurred postnatally but prior to hospital discharge, it is unclear how home BP assessment after discharge is relevant. Furthermore, if all of these complications occurred within the first 1-2 months postpartum, it may be inappropriate to conclude that home BP evaluation is associated with reduced complications for a year after delivery.

For clarification, adverse events that occurred postnatally but prior to hospital discharge are not included in the analyses. All study outcomes, including adverse events and costs, were measured post-discharge. This is now specified clearly in the methods in lines 130-131.
All of the adverse events described in Tables 2 and 3 occurred within the first 6 months following discharge. Across the intervention group and the two comparison cohorts there was only one additional adverse event between 6 and 12-months post-discharge. The observed 6-month difference did persist over 12 months. We have amended the text in the discussion to reflect that nuance. The text (lines 196-197) now reads "Although results were concentrated in the first 6 months following delivery, the patterns, including medical cost differences, persisted for 12 months." Additionally, lines 209 – 210 read "The greatest reduction occurred in the first six months, with only one adverse outcome in the second half of the first postpartum year" to further emphasize this point.

5. The term HELLP syndrome is often used a bit too liberally and often erroneously to describe a patient with preeclampsia with mild laboratory abnormalities. The most

common postpartum hypertensive complication that requires readmission is postpartum preeclampsia and this was not one of the adverse outcomes that was evaluated. Are the authors using the term HELLP syndrome synonymously with the term preeclampsia? If not, why wasn't readmission for preeclampsia studied?

Thank you for this comment. The authors agree that the term HELLP syndrome is often used for preeclampsia with mild laboratory abnormalities when the two entities are in fact different. In this study we are using an observed diagnosis of HELLP syndrome from medical claims. Although there is likely variability in how liberally or conservatively clinicians code for HELLP syndrome, we do not make a clinical judgment on the appropriateness of the diagnosis in this context and cannot control for mislabeling/misdiagnosis in patient records at time of encounter. The readmission outcome was all-cause readmission. We have amended to text to indicate that the readmission outcome is all-cause readmissions. The text (lines 132-135) now reads "Claims data were also used to assess total medical cost and healthcare service utilization including specialist visits, emergency department visits, cardiology visits, and all-cause inpatient readmissions in the first six months and first year following delivery." Lines 172-175 now state "Because the adverse events were concentrated in the first six months post discharge (only one event occurred after six months) we present utilization and cost results in the first 6 months of following discharge.""

In our study, we include all-cause readmission visits. However, 71% of the diagnoses for the observed ER visits and readmissions across the intervention and control groups were related to one of the post-partum adverse outcomes of interest. Acknowledging this important point that readmission for preeclampsia was not specifically evaluated in this study, a prior study that compared the remote monitoring program to traditional office based follow up did show a decrease hypertension related readmissions (Hirshberg et al, BMJ Qual Saf, 2018; reference 5). The fact that the majority of the diagnoses for the readmissions were related to the adverse outcomes is likely consistent with this prior finding.

6. Although there were less ER visits and hospital readmissions in the control groups, there is no information on the reason(s) for these visits. If these were visits for a gynecologic problem (e.g., miscarriage, STI) or a URI/pneumonia, for example, how would a postpartum home BP evaluation be relevant to these issues?

In our study, we include all-cause ER and all-cause readmission visits. Although we cannot guarantee that the potentially avoided ER visits and readmissions were related to a gynecologic problem, 71% of the diagnoses for the observed ER visits and readmissions across the intervention and control groups were related to one of the post-partum adverse outcomes of interest.

Since the complications studied are known sequelae of preeclampsia, albeit not always directly mediated by severe hypertension, it remains feasible that a postpartum home blood pressure monitoring program would be relevant to these outcomes. Also, the rate of non-OB/GYN admissions between groups was similar; 12% for the treatment group, 10% for the asynchronous controls, and 14% for the contemporaneous controls.

7. This study was conducted in the subjects and concurrent cohorts from 9.2017-4.2021. There is an established correlation between Covid19 infection and preeclampsia [AJOG Sept 2021;e1-17]. Do the authors have any information on the incidence of Covid19 in their

subjects and controls? Did the incidence of hypertensive disease increase in the final year of enrollment (table 1 shows higher incidence of hypertension in subjects and concurrent controls)? Were adverse postdelivery outcomes more common in women who had Covid during pregnancy? Do the authors feel that patients with Covid19 infection during pregnancy should be candidates for home BP evaluation?

We appreciate the reviewer highlighting the potential impact of Covid19 on preeclampsia, particularly in the treatment group and concurrent controls. Our data do not allow us to specifically tease out the impact of Covid19 on rates of hypertension. As this was claims-based data, unless a patient was hospitalized for COVID 19 or received antiviral treatment, we would not be able to accurately capture all infections. We do not have data on the last year of enrollment specifically, although general trends provided by the CDC suggest the incidence of chronic hypertension and hypertensive disorders of pregnancy has increased every year, even prior to the COVID pandemic.

In light of not having patient specific COVID-19 data, we cannot comment if adverse postdelivery outcomes were more common in women with COVID during pregnancy. However, we do not have any evidence to suggests that risks would be different for treatment and control group members or that COVID-19 impacted preeclampsia rates. Based on our institution data related to coronavirus, we found no association between SARS-CoV-2 infection and hypertensive disorders of pregnancy (Triebwasser et al, AJOG, Oct 2022) In a prospective cohort study of patients delivered between April 2020 and December 2020, seropositive patients were no more likely to be diagnosed with hypertensive disorders of pregnancy compared with seronegative patients. Additionally, there were no differences in severity of HDP by serostatus, nor in risk of HDP by severity of COVID, as well as no association in sensitivity analysis between seropositivity and preeclampsia SARS-CoV-2 infection by PCR was also not associated with hypertensive disorders of pregnancy. As such, we feel that patients with COVID19 infection during pregnancy are candidates for home blood pressure evaluation.

Reviewer #3:

GENERAL COMMENTS: The authors present a retrospective cohort study assessing the impact of a remote blood pressure monitoring program on adverse postpartum outcomes and medical costs. Patients enrolled in twice daily text-based blood pressure monitoring for ten days postpartum (n=1,700) were compared to two propensity score matched cohorts of i) patients at any of the three participating hospitals before remote monitoring program implementation (n=1,021) and ii) patients at other hospitals during the same time period as clinical use of the program (n=1,276). This is an important area of inquiry because postpartum hypertension is an important cause of morbidity and is pernicious because peak blood pressures occur 3 - 6 days postpartum when most patients are already home. Preliminary data from an RCT showed increased blood pressure ascertainment overall and reduced disparities with remote blood pressure monitoring, but the study was not powered to detect differences in outcomes. The results of this study showing that remote blood pressure monitoring reduced postpartum adverse outcomes, readmissions, ER visits and medical costs, and increased cardiology visits up to a year after delivery are very promising. A number of revisions listed below will further strengthen the manuscript.

SPECIFIC COMMENTS:

1. Inclusion criteria (patients with gestational hypertension, pre-eclampsia (PEC), chronic hypertension with superimposed PEC, HELLP syndrome, or eclampsia at the time of their delivery admission) are appropriate, but the exclusion of members who incurred over $12,500 in total medical costs per member per month (PMPM) in the prenatal period is not justified. These are likely the patients with comorbidities and therefore at the highest risks of adverse outcomes.

We thank the reviewer for this comment. As a robustness check we included members with >$12,500 PMPM in medical costs, rematched the cohorts, and reran the analyses. The results were qualitatively similar. Post-intervention medical costs were $120 PMPM lower for the treatment group compared to the asynchronous control and $97 PMPM lower compared to the contemporaneous control, though neither result was statistically significant. The lack of significance was a result of both substantially higher variance in observed costs and a reduced sample size. When we rematched, including the high-cost members the analytic sample size decreased the number of matches, thereby reducing the sample size and power of the analyses. Without including the high-cost outliers we report $32 lower PMPM lower relative to the asynchronous comparison group and $29 lower PMPM relative to the contemporaneous comparison group.

Given the nature of the inclusion criteria, everyone in this study is high risk and relatively high cost. We excluded the highest cost outliers to detect an effect on the more typical sick member not the highest of the high-risk due to noise in the data for these members. Our perspective is that excluding the highest cost members results would be more conservative and more generalizable. We have added this justification to the manuscript, which now reads "We also excluded patients who incurred over $12,500 in total medical costs per member per month (PMPM) in the prenatal period, as those outliers in the top 1% of cost do not represent the more typical patient and large variance in cost and potential regression to the mean could dilute estimates of the program effect on cost." (Lines 73-77).

2. The use of two propensity score-matched control cohorts is reasonable to account for confounders. An additional sensitivity analysis would be to perform a difference-in-difference analysis: i.e. compare change in outcomes before to after implementation of the remote blood pressure monitoring to the change in outcomes from the same before to after implementation timeframe in the hospitals where the program was never implemented. This would account for changes in practice over time.

We thank the reviewer for this suggestion. We have added the difference-in-difference robustness check to the robustness checks reported in Tables 3a and 4a of the Appendix. The Tables now report results from three different robustness checks; (1) A doubly robust estimator that that combines regression and propensity score methods to estimate the treatment effect, (2) Covariate-adjusted regressions on unmatched samples, and (3) Covariate-adjusted difference-in-difference using the matched sample. For each approach we estimated the standard errors using two methods, Robust SE using the robust sandwich formula, and the Bootstrap SE based on 1000 resamples. All specifications yielded results that were similar in direction and magnitude as those from the covariate-adjusted regressions on the propensity score matched sample we report in the text of the manuscript. Obviously, we did not run the DID specification for adverse outcomes because, by definition, adverse outcomes were only observed post-discharge.

3. The procedures for the propensity matching appear appropriate.

4. Outcomes were a composite of adverse outcomes identified through ICD-10 codes and healthcare resource utilization. Because not all diagnoses make it to claims documents there could have been under reporting of the adverse outcomes. This should be admitted as a limitation.

We appreciate the reviewer raising the issue of potential under diagnosis of adverse outcomes. Although this is a potential limitation, we have not called out that limitation in the text because for the following reasons. First, the potential for underdiagnosis of these serious adverse events is very low. It is highly unlikely that clinicians will miss coding a diagnosis of pulmonary edema or MI. Also, in cases where a patient presents with two adverse events (e.g., Eclampsia and stroke) and only one adverse event gets recorded our analyses would still include that patient as having an adverse event. If there were underdiagnosis of adverse events, we have no reason to believe it would occur at different rates between the treatment and control groups. Still, it is theoretically possible that there is underdiagnosis of adverse events and our data and analyses cannot preclude this possibility. Therefore, we have added this limitation in the discussion. Specifically, the text now reads "Another limitation is that it is likely that not all diagnoses make it to claims documents; there could have been underreporting of the adverse outcomes using ICD-10 codes and healthcare resource utilization" in lines 261-263.

5. None of the outcomes as designated the primary outcome. Although the issue of statistical power is moot because of the significant differences seen, it would be important to be explicit on how the investigators arrived at the sample size used.

The primary outcome was the reduction in adverse events. We have amended the text (lines 127-130) to read "Our primary outcome was having any prespecified adverse clinical outcome following delivery discharge, including stroke, DIC, eclampsia, pulmonary edema, Hemolysis Elevated Liver Enzymes Low Platelets (HELLP) syndrome, myocardial infarction, and cardiomyopathy."

The sample size for the intervention group was based on all eligible patients at the hospitals during the study timeframe who agreed to participate. As the program was implemented as standard of care in the hospitals during the study period, the sample size is fixed at all people meeting program inclusion criteria (hypertensive disorder of pregnancy, English speaking, access to text messaging). Our program data suggests that 99% of eligible patients consent to participation. The included patients are based on intention to treat; in other words, it includes patients who are enrolled who may not have sent in a blood pressure reading. However, as now mentioned in lines 94-96, data over the five years since implementation suggests that over 90% of patients text in at least one blood pressure and over 80% of patients send in multiple blood pressure readings over the ten day period.

6. Effect sizes, rather than only p-values, should be presented for the outcomes. And even after propensity score matching, adjusted analysis may still be performed as needed.

We thank the reviewer for the suggestion. We have added the effect sizes (Cohen's d) in the tables 4 and 5. We also clarify in the Methods section that results are from covariate-adjusted regressions. Specifically, we added the following language, "Differences on the extensive

margin, the likelihood of having any adverse event or any healthcare visit were assessed using logistic regression. Differences on the intensive margin, the number of adverse outcomes and healthcare visits were assessed using generalized linear models (GLM) with a log link function and negative binomial distribution. Cost outcomes were estimated using GLM with log link and a gamma distribution. Covariates for all regression models include age, risk score, comorbidities (hypertensive disorder, cardiovascular conditions, diabetes, asthma, obesity (BMI), anxiety disorders, Depression)." (Lines 136 - 142)

7. As disparities are an important feature of maternal morbidity and mortality data in the U.S., analysis should be performed by race/ethnicity to assess if remote blood pressure monitoring is equitable by reducing disparities in adverse outcomes and healthcare utilization.

We agree wholeheartedly with the importance of understanding the potential differential impact of programs like this one by race/ethnicity. Unfortunately, our race/ethnicity data are available only at the census block level rather than the individual level for all study members. We include census block level race/ethnicity in the matching process, but these data are not granular enough to make inferences about the program effect by race/ethnicity. For clarity, we have clarified in the text and in the note to Table 1 that race/ethnicity data are at the census block level, not the individual patient level.

Of note, we previously published data showing that remote blood pressure monitoring reduced disparities in blood pressure ascertainment (Reference 6). We hope that extrapolating from this data would suggest reduction in disparities in downstream effects of increased medication initiation and improved blood pressure control, although this current study cannot show that based on above limitations.

8. It must be acknowledged that while propensity score matching accounted for measured confounders there is still the possibility of residual confounding from unmeasured variables.

We thank the reviewer for this comment. We have added a sentence about this limitation in the manuscript; "Residuals bias from unobserved confounders remains a challenge in the study." (Lines 255- 256)

STATISTICAL EDITOR COMMENTS:

lines 29-33: First, the data provided relates to 6 months following delivery, not 1st year. See later comments re: tables 3, 4, 5, since many of the comparisons are NS. The third sentence is not correct, in that this was not a randomized study and therefore the differences cannot be generalized as necessarily applying to future cohort.

Thank you for this comment. The data provided emphasizes the 6 months following delivery but was collected over the first year. Only one additional adverse outcome occurred in the asynchronous group in months 7-12 after delivery. All other outcomes occurred in the first six months. Although this specific study was not a randomized study, the impact of the study on

short term outcomes is supported by a prior randomized trial (reference 5), thereby supporting our concluding statement in the third sentence.
The study examined adverse event rates and utilization and cost outcomes over 12 months, though the impact of the intervention on adverse events was concentrated in the first 6 month. We thank the reviewer for highlighting the potential confusion around the timeframes for the results. We have made changes throughout the manuscript to clarify that the results we present are those observed in the first 6-months after discharge (Lines 172-175 now state "Because the adverse events were concentrated in the first six months post discharge (only one event occurred after six months) we present utilization and cost results in the first 6 months of following discharge"), but" the results persist over the 12-month period" (Lines 198-199)
Regarding the generalizability of findings, we have revised the abstract to be more specific with regard to the population to which the results may generalize. Specifically, we say "…broad implementation of evidence-based remote monitoring programs may reduce postpartum adverse outcomes, thereby reducing morbidity and mortality in populations like the one studied here." (Lines 29-31)
However, we disagree that our study lacks generalizability. The advantage of randomized studies is they produce producing accurate effect estimates (high internal validity). Propensity score matched studies like ours try to address confounders that can bias effect estimates and reduce internal validity. However, both experimental designs (e.g., RCTs) and quasi-experimental designs (e.g., PSM studies) have the same limitations in terms of generalizing results beyond populations similar to the ones studied. (See Campbell, Donald T., and Julian C. Stanley. Experimental and quasi-experimental designs for research. Ravenio books, 2015.)

lines 80-81: Was enrollment voluntary and if so, how many were enrolled and how many declined or did not complete the "exposure"? Should include a flow diagram.
Enrollment in the program in voluntary, although the program was standard of care at the hospitals used for the study. Based on prior experience, over 99% of patients who are eligible enroll. We do not have individual level data on program decline rates or those not completing the "exposure" for the specific Independence members used for this study. However, our internal program dashboard suggests that over 90% of patients send in at least one blood pressure over the 10 day period, with over 80% of patients sending in multiple blood pressures. This compliance rate is closer to 95% at the sites with more Independence members. As we do not have this level of detail for the individual members who met eligibility criteria, we are unable to provide a flow diagram. However, the general compliance is mentioned in the methods in lines 94-96.

lines 89-90: The emulation of randomization only applied to known confounders. Since the groups were not randomly allocated, unknown confounders remain unknown.
We agree with the reviewer's point and have called out this limitation in the discussion with the following text "Residuals bias from unobserved confounders remains a challenge in the study." (lines 255-256).

lines 94-96: Need to provide a reference or supplemental material demonstrating the risk score calculation.
We do not have access to the risk score calculation. We used Verisk Health's proprietary

diagnostic cost group (DxCG) risk score, a commonly used risk-adjustment measure for cost data. (E.g., Wagner, T. H., Upadhyay, A., Cowgill, E., Stefos, T., Moran, E., Asch, S. M., & Almenoff, P. (2016). Risk adjustment tools for learning health systems: A comparison of DxCG and CMS-HCC V21. Health Services Research, 51(5), 2002–2019. https://doi.org/10.1111/1475-6773.12454).
We have amended the text to clarify the source of the risk score. Specifically, lines 108-112 now read "We matched the two comparison groups to the treatment group one-to-one without replacement on the following measured pretreatment variables: age, race, ethnicity, delivery month, prenatal total medical cost, Diagnostic Cost Group (DxCG) risk score, and preexisting chronic conditions (hypertension, diabetes, cardiovascular conditions, depression, and anxiety disorder).

Methods: Need to explain more fully the treatment group BP monitoring, its compliance, frequency, loss to follow-up etc.
Thank you for this comment. More details for the treatment group are discussed in lines 93-96 of the methods, including frequency and compliance.

Should include a Table similar to Table 1 but comparing the matched cohorts.

Table 2: For this comparison, the rates (2.9% vs 4.7%) are significantly different, but only at a p = 0.04 level. Should include the CIs for each proportion, which are 2.0-4.2% and 3.5-6.2%, respectively.

Table 3: For the contemporaneous cohorts, the difference in rates is NS. They were 3.2% (2.3-4.4%) vs 4.5% (3.4-5.8%), with p = 0.11. Applying other test statistics, Chi-square = 2.74 or OR = 1.41 (0.94-2.12), both of course with same p = 0.10

We thank the reviewer for highlighting the potential confusion with how the results were presented in the Tables. The p-values presented in Tables 2 and 3 were 1-tailed results from a t-test of proportions where the p-values are 0.02 and 0.05, respectively, but were unadjusted. We have removed significance tests from Tables 2 and 3 and focus on the regression adjusted results presented in Tables 4 and 5, which we have updated to include the ORs and CIs for the likelihood of having any adverse event in the first 6 months post-discharge (extensive margin) and effect sizes (Cohen's d) for the results of GLM regressions comparing the number of adverse events and healthcare visits (intensive margin).

Table 4: Should change the format to show the actual counts in each matched group as n (%), rather than as rates per 1000. Should include CIs for the differences and omit the column of p-values. For these cohorts, none of the numerical differences are statistically significant. For those with relatively low rates (ER visits or readmits), the results are both NS and vastly underpowered to discern such small differences.

Table 5: Again, should change the format, include CIs and omit the p-values. The results

for these contemporaneous groups are each NS.

We thank the reviewer for their perspective on the presentation of the results. In tables 4 and 5 we now present results as regression-adjusted counts of events along with 95% CI's and effect sizes (intensive margin) and OR with 95% CI's and effect sizes for having any adverse outcome or healthcare utilization event (extensive margin).

Should include the covariate balance figures to main text. Suggest omitting the K-S stats comparisons and the remaining text explaining.
Thanks for the suggestions. We have moved the covariate balance figures to the main text (Figure 1 for the asynchronous cohort and Figure 2 for the contemporaneous cohort) and omitted the K-S statistics.

Editors Comments:
1. Please carefully review the comments by the Statistical Editor regarding the statistical analysis of the included data.

Thank you for this suggestion. Please refer to above responses regarding the statistical analysis.

2. Please provide additional data on the intervention in the Methods section including how the intervention was implemented, whether it was voluntary, and any additional detail that may be helpful to readers.

Thank you for these points. We have provided additional data regarding the intervention in lines 78-96.

3. Percentages should be included in tables 2-5.

We appreciate this suggestion. Percentages are now included in the Tables 2-5.

4. As one of the Reviewers points out, given that this intervention represents in essence a "natural experiment" based on an insurance-based intervention that study would seem to be well suited to a difference in difference analysis.

Thank you for this suggestion. In response to this comment and Reviewer 3, Comment 2,we have added the difference-in-difference robustness check to the robustness checks reported in Tables 3a and 4a of the Appendix. The Tables now report results from three different robustness checks; (1) A doubly robust estimator that that combines regression and propensity score methods to estimate the treatment effect, (2) Covariate-adjusted regressions on unmatched samples, and (3) Covariate-adjusted difference-in-difference using the matched sample. For each approach we estimated the standard errors using two methods, Robust SE using the robust sandwich formula, and the Bootstrap SE based on 1000 resamples. All specifications yielded results that were similar in direction and magnitude as those from the covariate-adjusted regressions on the propensity score matched sample we report in the text of the manuscript. Obviously, we did not run the DID specification for adverse outcomes because, by definition, adverse outcomes were

only observed post-discharge. We reference the robustness checks in the methods in lines 142-146 and in the results in lines 191-192.

Date: 02/02/2023
To: "Adi Hirshberg" [redacted]
From: "The Green Journal" em@greenjournal.org
Subject: Your Submission ONG-22-1756R1

RE: Manuscript Number ONG-22-1756R1

A remote blood pressure monitoring program's impact on postpartum adverse outcomes

Dear Dr. Hirshberg:

Thank you for your submission. Your manuscript revisions were reviewed by the Editors and there remain additional questions that were raised regarding the manuscript. These comments are outlined below. Please provide a point-by-point response to each of these comments.

The revised manuscript should indicate the position of all changes made. Please use the "track changes" feature in your document (do not use strikethrough or underline formatting).

Your submission will be maintained in active status for 21 days from the date of this letter. If we have not heard from you by 02/16/2023, we will assume you wish to withdraw the manuscript from further consideration.

EDITORS' COMMENTS (for submitted R1 Manuscript)

1. General: The Authors have outlined and shown the results of matching for the asynchronous and concordant groups, which resulted in excellent matching for all the given variables. However, between the original version and the revised version, the counts for the adverse outcomes have been modified. The primary outcome remains the same (lines 139-142), so there is no justification for changing the counts from the original ones: 30 vs 48 in asynchronous (each with N = 1021) and 41 vs 57 in the concordant (each with N = 1276). The resultant primary outcome ORs are statistically significant for the asynchronous, but not for the concordant matched cohort comparisons.

2. Regarding costs, if these patients were individually identified, then why were health care costs estimated, rather than simply summed from the record? That would eliminate the need for truncating some of the data and would directly compare the real-time costs of the two treatment arms.

3. Lines 29-33: The comments in Abstract should only reflect the study at hand, whether another study was an RCT is not relevant to this Abstract.

4. Lines 80-81: How many n (%) of patients in each of these cohorts (1) enrolled and (2) provided at least one BP measurement during the period of this study?

5. Lines 94-96: Need to include a reference for the use of DxCG risk score.

6. Methods Section:

a. re: BP monitoring, frequency etc: Need to provide information specific to these cohorts during the time of this study.

b. Tables 2 and 3: We must insist that the differences in adverse event rates be presented in standard fashion so that they would be understood by our readers. The usual methods for calculation of CIs and resultant p-values should be adhered to. There is no justification for a priori application of one-tailed t-test. There is no justification for assuming that the treatment arm can only have better results than the control group. As such, the differences in the primary outcome (any adverse event) are NS different for the contemporaneous group, but are statistically significant for the asynchronous group. If the same format as in the original Tables 2 and 3 are used, then the differences must include appropriate CIs or by application of Fisher's or Chi-square testing, which demonstrates that all row entries for differences, except for the asynchronous comparison of "Any" are NS.

c. Regarding the updating of the adverse event counts: The description on lines 109-112 of the original submission or on lines 127-130 of the 1st revision specify which adverse outcomes to be considered as the primary. Nothing in either description states that any adjustment for or expansion of the adverse outcome counts was to be implemented. Need to adhere to the original definition of the primary outcome and omit the intensive and extensive margins.

7. Finally, this is promising work on an important topic, but the results need to be presented in standard format for our readers. Fortunately, the counts for adverse outcomes are few, but that limits statistical power to discern differences.

--
Sincerely,
Jason D. Wright, MD
Editor-in-Chief

Thomas Riggs, MD, PhD
Deputy Editor, Statistics

The Editors of Obstetrics & Gynecology

_____

Adi Hirshberg, MD

Department of Maternal Fetal Medicine

Hospital of the University of Pennsylvania

3400 Spruce Street-2nd Floor Courtyard

Philadelphia, PA 19104

[redacted]

[redacted]

February 21, 2023

Obstetrics and Gynecology Editorial Staff:

Thank you for the opportunity to review and resubmit second revisions to our manuscript entitled **"A remote blood pressure monitoring program's impact on postpartum adverse outcomes**". As previously stated, all of the authors listed meet the authorship requirements and no authors have any conflicts of interest or financial disclosures to report. Persons acknowledged provided permission to be named. The manuscript has not previously been submitted to *Obstetrics and Gynecology* and is solely submitted to *Obstetrics and Gynecology.* It is not currently in consideration for publication with another journal and will not be submitted elsewhere unless a final negative decision is made. This work was presented at the Academy Health 2022 Annual Research Meeting, Washington DC, June 4-7, as an oral presentation.

We believe that our revisions will satisfactorily address all reviewer concerns and hope that you find it acceptable for publication in *Obstetrics and Gynecology.*

Sincerely,

Adi Hirshberg, MD

[redacted]

EDITORS' COMMENTS (for submitted R1 Manuscript)

Comment 1. General: The Authors have outlined and shown the results of matching for the asynchronous and concordant groups, which resulted in excellent matching for all the given variables. However, between the original version and the revised version, the counts for the adverse outcomes have been modified. The primary outcome remains the same (lines 139-142), so there is no justification for changing the counts from the original ones: 30 vs 48 in asynchronous (each with N = 1021) and 41 vs 57 in the concordant (each with N = 1276). The resultant primary outcome ORs are statistically significant for the asynchronous, but not for the concordant matched cohort comparisons.

Response 1: We thank the reviewer for the comment and want to be clear about the calculation and presentation of the adverse event outcome.

In the original submission we compared the simple counts of number of patients within each group who had any adverse event using a chi-square analysis. However, the first round of reviewer comments caused us to look more closely at our analytic approach. For example, one reviewer suggested we run a difference-in-differences model to estimate the impact of the intervention on changes in health care utilization and costs. We took those suggestions to heart and conducted a full suite of robustness checks across all of our outcomes, which has ultimately made this a much stronger paper.

We believe that using logistic regression to examine the impact of the program on the probability of having an adverse event is a more methodologically appropriate test. The logistic model provides an estimate of the number of events in each group controlling for some of the remaining differences between the groups that exist even after matching. The odds ratio allows us to quantify the reduction in the probability of having an adverse event and, importantly, to make inferences about what the expected impact would be in other similar trials/programs in the future. Also, after conducting numerous robustness checks (8 different model specifications x 2 comparison groups) we found that the program was associated with significantly fewer adverse events in 15 of the 16 comparisons. Logistic regression is the preferred specification because it is a more robust and appropriate test and because it more accurately represents the preponderance of the evidence regarding the impact of the program on having an adverse clinical outcome.

Nevertheless, we agree that the chi-square test provides some useful descriptive information about the observed events and have reincorporated this test into the manuscript in addition to the results from the logistic regression model.

Specifically, in the abstract we state "*In the first six months following delivery patients enrolled in remote monitoring were less likely to have the composite adverse outcome than those in the asynchronous cohort (2.9% vs 4.7%; OR 0.61, 95% CI 0.40 – 0.98) and trended lower relative to the contemporaneous cohort (3.2% vs 4.5%; OR 0.71, 95% CI 0.47 – 1.07)* . [Lines 20-24]

In the main text of the manuscript, we state "*Chi-square analysis showed that significantly fewer program participants had an adverse clinical outcome relative to the asynchronous cohort (2.9% vs 4.7%; OR 0.61, 95% CI 0.40 – 0.98) and trended lower relative to the contemporaneous cohort (3.2% vs 4.5%; OR 0.71, 95% CI 0.47 – 1.07) but was not statistically significant. Multivariable logistic regression analyses, controlling for confounders, also show that patients enrolled in the remote blood pressure monitoring program were less likely to experience any adverse clinical outcome compared to patients in the*

asynchronous cohort (2.3% vs 4.5%; OR 0.54, 95% CI 0.33 – 0.87) and the contemporaneous cohort (2.9% vs 4.9%; OR 0.59, 95% CI 0.40 – 0.88).” [Lines 172-184]

We have focused our results to report the chi square analyses in the text and table and mention the multivariable logistic regression in the results. The tables reflect the chi square analyses. If the editors would like us to present the logistic regression results in the table instead, we are happy to do that.

Comment 2: **Regarding costs, if these patients were individually identified, then why were health care costs estimated, rather than simply summed from the record? That would eliminate the need for truncating some of the data and would directly compare the real-time costs of the two treatment arms.**

Response 2: We thank the reviewer for the comment and apologize for any confusion. For clarity, all outcomes (i.e., adverse events, healthcare utilization outcomes, and medical costs) are not estimated but are observed in real-time in the medical claims data, as correctly point out. As is typical in studies using medical claims, direct comparisons of observed costs are conducted using regression models in which costs are regressed on treatment (0,1) and select covariates. The resulting coefficient for the treatment variable is the direct comparison of the cost for those in the treatment compared to those who did not get the treatment. The coefficient is referred to as an estimate, or point estimate, with an associated confidence interval around the difference. So, you are correct, the costs are observed not estimated. For clarity we have amended the text to read “*A log link and a gamma distribution was used to test differences in total incurred medical costs.”* [Lines 146-147]

Comment 3. **Lines 29-33: The comments in Abstract should only reflect the study at hand, whether another study was an RCT is not relevant to this Abstract.**

Response 3: We agree on this point. There are no references to or suggestions of an RCT in the abstract and all methods and results relate to the study at hand. The RCT is only referenced in the introduction when discussing findings from prior studies of remote monitoring.

Comment 4. **Lines 80-81: How many n (%) of patients in each of these cohorts (1) enrolled and (2) provided at least one BP measurement during the period of this study?**

Response 4: Thank you for this comment. All patients in the program were enrolled, as this was standard post-delivery discharge follow-up care at the hospitals included. By cohort definition, no patients in the asynchronous or synchronous cohort were enrolled. For the purpose of this study, we do not have individual patient level data as to how many patients provided at least one blood pressure measurement during their monitoring period. However, as included in the revised manuscript based on previous reviewer comments, historical data, gathered over five years of implementation and deployment as standard of care, suggests that 90% of patients text in at least one blood pressure and over 80% of patients send in multiple blood pressure readings over the ten-day period. While we do not know if the patients in the program cohort definitively sent in at least one blood pressure, our analysis was performed as intention to treat. That being said, we have dashboards showing continued compliance, with higher engagement in hospitals with higher prevalence of commercially insured patients, suggesting the overwhelming majority of patients included in the program cohort would have sent in at least one blood pressure. We have included this as a limitation, stating “*Lastly, we do not have individual level data regarding patient adherence to blood pressure monitoring for those enrolled in the program and therefore cannot know with certainty the effect on adverse outcomes; however, based on historical data, compliance is high, at over 90%, suggesting most patients do in fact send in blood pressure data during the needed time period*”. [Lines 279-282]

Comment 5. **Lines 94-96: Need to include a reference for the use of DxCG risk score.**

Response 5: We have added a reference for the DxCG risk score, now reference 10.

The evolution of DxCG, the gold standard in risk adjustment and predictive modeling. Cotiviti. Accessed February 10, 2022. https://resources.cotiviti.com/population-health-analytics/cotiviti-whitepaper-evolutionofdxcg

Comment 6. Methods Section:

Comment 6a. **re: BP monitoring, frequency etc: Need to provide information specific to these cohorts during the time of this study.**

Response 6a: Thank you for this comment. Details of the program cohort are provided in the methods based on feedback after the initial submission. For clarity, patients received twice daily requests for blood pressure readings for ten days following discharge. This is the standard of care for how the program is implemented based on prior work and a prior RCT evaluating the program. Patients in the asynchronous or contemporaneous cohort received blood pressure follow-up at the discretion of their hospital/physician. This has been clarified in lines 108-109. We do not have this level of data available but given the novelty of our program, ACOG baseline recommendations, and the current state of postpartum blood pressure monitoring across the country, we suspect the majority of patients in the other cohorts likely had only one blood pressure check in the 2 weeks after delivery.

Comment 6b. **Tables 2 and 3: We must insist that the differences in adverse event rates be presented in standard fashion so that they would be understood by our readers. The usual methods for calculation of CIs and resultant p-values should be adhered to. There is no justification for a priori application of one-tailed t-test. There is no justification for assuming that the treatment arm can only have better results than the control group. As such, the differences in the primary outcome (any adverse event) are NS different for the contemporaneous group but are statistically significant for the asynchronous group. If the same format as in the original Tables 2 and 3 are used, then the differences must include appropriate CIs or by application of Fisher's or Chi-square testing, which demonstrates that all row entries for differences, except for the asynchronous comparison of "Any" are NS.**

Response 6b: We thank the reviewer for this comment. Tables 2 and 3 were included to provide additional context and detail related to the specific adverse events we included in our composite measure of adverse clinical outcomes. We have clarified the result of the primary outcome by setting the first line to portray the results of our primary outcome using Chi-square testing, with p values, OR, and CI. The primary outcome is the dichotomous composite outcome of any of the 8 adverse clinical outcomes. We removed the individual tests in our previous revision because our outcome of interest is not the reduction of any particular adverse event but rather the probability of having the composite adverse outcome which includes any adverse clinical outcome across all 8 categories. We were not powered to find significant differences for any of the 8 individual outcomes and we had no a-priori hypotheses about which ones should be more or less impacted by the program. Providing tests across the individual outcomes would potentially be misleading with regard to the intent and design of the study. In this revision we also omit the individual tests and focus more directly on the primary outcome of having an any of the adverse events. The dichotomous composite adverse outcome has always been the a priori outcome of the study as we hypothesize that this postpartum blood pressure monitoring program may help reduce any of these hypertension related adverse events. As described in our response to Comment 1, we include the chi-

square test in the text of the manuscript as well as the results from the logistic regression model, which is a more robust test and a more appropriate test of our hypothesis.

Also, in response to an earlier reviewer comment, we do note that "*Although we were not powered to detect program effects within each of the eight clinical categories, we did observe fewer postnatal adverse outcomes in the intervention group in each of the eight clinical categories.*" [Lines 218-220]

Finally, there are no references to one-tailed tests in the manuscript.

Comment 6c. **Regarding the updating of the adverse event counts: The description on lines 109-112 of the original submission or on lines 127-130 of the 1st revision specify which adverse outcomes to be considered as the primary. Nothing in either description states that any adjustment for or expansion of the adverse outcome counts was to be implemented. Need to adhere to the original definition of the primary outcome and omit the intensive and extensive margins.**

Response 6c: We appreciate the call for clarity regarding the measured outcomes. As we state in both the original submission and the first revision "*The purpose of the current study was to evaluate the impact of a text-message based remote blood pressure monitoring program on adverse postpartum clinical outcomes and costs in patients with a hypertensive disorder of pregnancy. We hypothesized that program enrollment would result in decreased adverse outcomes and healthcare costs due to the potential for early intervention and improved transitions of care compared to two matched control groups*." [Lines 59-64]

One way to think about reducing adverse postpartum clinical outcomes is to look at the number of patients who experience any adverse outcome. Another equally valid way to think about it is a reduction in the total number of adverse events, as some women experience multiple events. There is reason to expect that the program would positively impact both ways of measuring adverse events. In fact, the total number of adverse events is more closely aligned with the patient's total morbidity, and importantly with the amount and type of healthcare services they consume and resultant medical costs, which are also important outcomes of this study (Tables 4 and 5).

As you correctly point out, we state in the methods section that the primary outcome is "*having any prespecified adverse clinical outcome following delivery discharge*." In this revision we further clarify our measure of the outcome variable by amending the text to read "*Our primary outcome is* a *composite measure of having any prespecified adverse clinical outcome following delivery discharge*." [Lines 131-132] We also make it clear that tests of this primary outcome include multivariable logistic regression for the probability of experiencing any adverse clinical outcomes and the more descriptive chi-square test.

We also clarify that our secondary outcomes include the number of healthcare utilization events and total incurred medical costs, that accrue across all observed adverse clinical outcomes. For clarity and consistency with the primary intent of the study, we have amended the text to explicitly state "*Claims data were also used to assess our secondary outcomes including the medical cost and healthcare service utilization including specialist visits, emergency department visits, cardiology visits, and all-cause inpatient readmissions in the first six months following delivery."* [Lines 136-139]

Comment 7. Finally, this is promising work on an important topic, but the results need to be presented in standard format for our readers. Fortunately, the counts for adverse outcomes are

few, but that limits statistical power to discern differences.

Response: We agree that the remote monitoring program is important work and are grateful for the opportunity to describe its impact on reducing adverse clinical outcomes for high-risk patients, improving appropriate specialist care, and reducing avoidable and costly visits to the ER and inpatient admissions.

As we described in our response to item 6b above, we agree with your assessment that the individual outcomes are too rare to detect, and it would be misleading to describe or present individual pair-wise tests for each event. Our hypothesis was that the program would reduce the likelihood of having any adverse event across the 8 categories relative to both control groups. The presentation of results are now more clearly focused on the primary outcome of having any adverse event and we have clarified the definition of the primary outcome as being a "*composite measure of having any prespecified adverse clinical outcome following delivery discharge*." [Line 132]