NBER WORKING PAPER SERIES

HOW DO PHYSICIANS RESPOND TO NEW MEDICAL RESEARCH?

Philip DeCicca
Maripier Isabelle
Natalie Malak

Working Paper 32656
http://www.nber.org/papers/w32656

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2024

We thank participants from the American Economic Association meetings, the Canadian Health Economics Study Group meetings, the Societe Canadienne de Science Economique meetings, as well as seminar participants from Universite Laval and the University of Alabama in Huntsville, and anonymous referees for helpful comments. We would also like to thank Ronen Avraham and Diane Alexander for graciously making their respective datasets publicly available. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How Do Physicians Respond to New Medical Research?
Philip DeCicca, Maripier Isabelle, and Natalie Malak
NBER Working Paper No. 32656
July 2024
JEL No. I11

ABSTRACT

What happens when the findings of a prominent medical study are overturned? Using a medical trial on breech births, we estimate the effect of the reversal of such a medical study on physician choices and infant health outcomes. Using the United States Birth Certificate Records from 1995-2010, we employ a difference-in-differences estimator for C-sections, low Apgar, and low birth weight measures. We find that the reversal of a multi-site, high profile, randomized control trial on the appropriate delivery of term breech births, the Term Breech Trial (TBT), led to a 15-23 percent decline in C-sections for such births at a time when the overall trend in C-sections was rising. We find our largest estimated effects amongst traditionally disadvantaged groups. However, we do not find that such a change in practice had significant impacts on infant health. Contrary to prior studies, we find that physicians updated their beliefs quickly, and do indeed adjust to new medical research, particularly young physicians, prior to mandatory policy or professional guidelines.

Philip DeCicca
Ball State University
117 Whitninger Building
Muncie, IN 47303
and NBER
ppdecicca@bsu.edu

Maripier Isabelle
Department of Economics
Laval University
Quebec City, Queb G1V 0A6
Canada
maripier.isabelle@ecn.ulaval.ca

Natalie Malak
University of Alabama in Huntsville
301 Sparkman Drive,
Huntsville, AL 35899
npm0003@uah.edu

1. Introduction:

This paper examines how the public reversal of a prominent medical study's results influences physician procedural choice. While many studies have examined the impact that novel medical findings can have on practice patterns (c.f. Howard et al., 2017; Oster, 2018; Phelps, 1992; Price and Simon, 2009), we are--to our knowledge--the first to investigate if the overturning of an initial high-profile medical trial can lead physicians to change their behaviour, even when it means going against well-established beliefs. We focus on an influential multi-site randomized control trial on the appropriate delivery of term breech[1] births, which began to be challenged in the medical literature four years after its publication. In contrast to studies showing obstetricians' responses are generally slow, but faster for relatively "advantaged" patients (ie. high education) (c.f., Oster, 2018; Price and Simon, 2009), our results seem to imply the opposite. In particular, we find that the overturning of the initial high-profile study had arguably large and immediate impacts on provider behaviour, particularly that of younger providers, and that traditionally disadvantaged patients (ie. non-white, and minimal education) were most impacted. In line with the debate in the literature regarding the impact of elective C-sections on infant health, we find mixed evidence on the potential long-run consequences on newborns' outcomes.

Our work also relates to existing research on how physicians respond to changes in medical news. There are several articles on how physicians respond to news by changing procedural mix, which correspond to the various mechanisms through which information diffuses into physician practice, e.g., behavioral factors (Staats et al., 2018), peer effects (Berez et al., 2018), organizational incentives (Howard et al., 2017), and physician skill (Wu & David 2022).

[1] Breech babies are defined as babies that have not turned head down (vertex position) in the womb by term (37 weeks of gestation). The likelihood of being in breech position is greater among small fetal sized and premature babies since they have more room to move around and change positions (Roberts et al., 1999). We thus limit the sample to term births, which will be discussed in the Methods section and Appendix A.

Substantial literature has found physician procedural choice (ie. practice style) can be affected by financial incentives (Almond and Doyle, 2011; Almond et al., 2010; Currie et al., 1995; Gruber et al., 1999), malpractice liability (Currie and McLeod, 2008; Kessler and McClellan, 1996), and patient knowledge/characteristics (Currie et al., 2010). Of particular note, Epstein and Nicholson (2009) conclude that although physicians may respond to changes in clinical guidelines, obstetricians are not likely to converge over time to community standards and they do not substantially revise their prior beliefs regarding the use of C-sections. Their interpretation is that "a considerable amount of practice variation is due to idiosyncratic physician perceptions regarding the appropriateness of specific treatments". Similarly, Cutler et al. (2019) conclude that variations in procedure use are driven by physicians' beliefs about appropriate treatment that need not be supported by clinical evidence. We find that the initial publication did not lead to changes in C-section rates; however, our findings on the reversal suggest some obstetricians are indeed responsive to cutting edge information.

The "Term Breech Trial" (TBT hereafter) was published in October 2000, and saw its initial findings overturned starting in late 2004. The initial results from the multi-center, multi-country trial concluded that full term singleton breech babies delivered via C-sections had better infant health outcomes than vaginal deliveries (Hannah et al, 2000). Shortly after, many different professional organizations, including the American College of Obstetricians and Gynaecologists (ACOG), codified these findings by putting forth guidelines that recommended a woman should have a C-section when giving birth to a full-term singleton breech baby (ACOG, 2001).[2] Recent research focusing on the causal impact of the trial, however, shows that the initial findings from

[2] The study had global reach, and its influence was documented in several countries, including Canada (Baker et al 2022), Sweden (Alexandersson, 2005), Australia and New Zealand (Sullivan et al., 2009), the Netherlands (Rietberg et al., 2005), and Denmark (Jensen and Wüst, 2015).

the TBT might have had little to no effect on the rate of C-sections performed on term breech babies in certain contexts. This is not surprising given that a breech position is one of the highest risk factors for C-section deliveries (see Currie and MacLeod, 2017).

As noted, the initial TBT findings determined the standard of care, but only for a few years. In late 2004, follow-up studies of the TBT showed that the initial findings were partly driven by the features of the trial (including the selection of candidates) not being representative of typical practice of vaginal breech births. When taking physician expertise and risk selection into account, among other factors, no significant differences in birth outcomes between vaginal and C-section delivery for full-term singleton births emerged (Kotaska, 2004).[3] In effect, the conclusions from the initial TBT results were reversed; the evidence strongly suggested that all else equal, C-section deliveries generally did not lead to superior outcomes for the newborn.

Focusing on U.S. physicians' choice of procedure for term breech births, we first document the causal impact of the TBT in the years following its publication in late 2000. Similar to Jensen and Wüst (2015) who analyze data from Denmark, we observe no change in the C-section rate for first time mothers after the initial TBT results. Unlike them, however, we also find no evidence of an increase in C-sections for higher-parity term breech births. This is perhaps not too surprising given that in the U.S. the C-section rate for breech births was very high by international standards: around 80 percent *prior* to the release of the TBT findings. While one might expect a small response from the overturning of the TBT results in 2004, especially in the U.S. where we find no response to the initial trial, we find a large drop in C-sections for breech births at a time when the overall trend in C-section use was moving upward. In particular, we find consistent evidence that the reversal of the original TBT findings led to a decline between 15-23 percent in the use of C-

[3] Other studies have also documented many concerns with the TBT and the external validity of its findings. See, for example, Alarab (2004), Turner (2006), Goffinet et al. (2006), and Glezerman (2006).

sections during term breech deliveries. Moreover, we find stronger results in counties with younger physicians and more physicians who studied abroad, but no differences by physician gender. We believe that this paper contributes important evidence on the impact that debates within the medical and research communities ultimately may have on physician choices, while documenting that physicians respond to research evidence prior to mandatory policy or professional guidelines.

2. Methods:

Using the United States Birth Certificate Records from 1995-2010 for all states and the District of Columbia, we model the impact of the initial TBT findings and then their reversal separately, employing a difference-in-differences estimator. However, we do not employ a standard difference-in-differences estimation comparing groups with different regulations, but instead different types of births under different knowledge sets. Our control group consists of complication-free births while the treatment group consists of breech births. Table 1 displays all summary statistics separated by policy period and treatment group (breech versus vertex births) for our preferred sample.[4] Our preferred sample is composed of 28,060,177 second or higher parity full-term singleton births, for which the mother did not have a previous C-section or any important pregnancy risk factors. Jensen and Wüst (2015) note that physicians were more likely to respond to the publication of the TBT when delivering higher parity births. Moreover, risky pregnancies and those following a previous C-section are associated with higher rates of C-sections. We therefore focus on a sample less likely to see variations in its C-section rates based on its composition rather than on the publication of new evidence on term breech births.

[4] See Appendix A for in depth discussion of our data.

For transparency and sensitivity reasons, we create four different treatment-control groups: A, B, C, and D (our preferred sample). In Group A, the treatment group consists of all full-term breech singleton births. We start with this grouping as our baseline sample as it consists of the group directly targeted by the TBT. However, it is well-documented that women with prior C-sections usually receive subsequent C-sections, these births should be less responsive to changes in guidelines, and presents a mechanical increase in C-section rates. Group B thus removes all mothers who have had a previous C-section from our baseline group, Group A. Given that Jensen and Wüst (2015) found no impact of the TBT trial for first-born breech babies in Denmark, we create Group C which removes first-birth parities from Group B. This allows us to understand if there is a difference in the impact of the TBT and its reversal based on birth parity in the US. Group D represents our final sample, and also removes "risky" pregnancies (pregnancies experiencing chronic or gestational diabetes, blood pressure, or eclampsia) from Group C. We define Group D narrowly with the intention of removing other factors that would result in a physician leaning towards performing a C-section – as a result, our interpretation for the results from this group can be thought of as the impact of new evidence on the safety of delivery methods for breech births on physicians' decisions for births that are most likely to be marginal cases.

We estimate the following linear probability model:[5,6]

$$(1)\ A_{it} = \beta_0 + \beta_1(Breech_{it} x\, TBT_t) + \beta_2 Breech_{it} + \beta_3 X_{it} + \beta_4 P_{it} + \beta_5 S_{st} + \beta_6 COUNTY + \beta_7 YEAR + \beta_8 MONTH + \varepsilon_{it}$$

Here, A_{it} represents whether individual i in year t gave birth via C-section, which is our main outcome. We also use a similar specification to investigate whether or not the newborn received a

[5] Models are also estimated via probit and yield very similar implied effects. See Appendix F.

[6] In Appendix G we use equation (1) but look at annual treatment effects instead of the overall effect. The annual treatment effects show a consistently statistically significant impact over the treated time period for the reversal of the TBT on C-section and low birth weight outcomes.

low 5-minute Apgar score (defined as less than nine, out of a maximum score of ten) or had a low birth weight (less than 2500 grams). Looking at the impact of the TBT itself, our main variable of interest, $Breech_{it}x\ TBT_t$, is the interaction of whether a given delivery/birth was breech during the period which corresponds to the TBT initial results being in force (November 2000- October 2004). We redefine this variable when looking at the initial TBT findings being overturned, to be the interaction of a breech indicator and the period between November 2004 and December 2010 (the end of our sample).[7] When focusing on this second event, we remove observations from the period spanning from November 2000 to October 2004, when the initial results from the TBT were operative. Our results, however, remain mostly unchanged when re-estimating the same model on the full period (January 1995 to December 2010) covered by our data.

X_{it} and P_{it} represent characteristics of mothers and their pregnancies at the individual-year level, respectively. S_{st} includes controls for the annual state Medicaid fee difference between C-section and vaginal deliveries, common medical malpractice tort reforms, as well as annual county unemployment rates for reasons discussed in Appendix A. All models include county fixed effects, $COUNTY$, to account for potential geographic heterogeneity in the utilization of C-sections that may be due to plausibly time-invariant factors such as access to medical services or local physician practice patterns (Currie and MacLeod, 2008; Newhouse and Garber, 2013). We also include year fixed effects, $YEAR$, to capture any secular trend that is shared across places, as well as month fixed effects, $MONTH$, to account for potential seasonality in C-section births.[8,9] Lastly, we cluster

[7] The initial TBT results were published at the end of October 2000, and the overturned TBT results were published at the end of October 2004. Subsequently, the ACOG adjusted their guidelines on breech deliveries in December 2001 and July 2006, well over a year from both the initial and reversed TBT studies.
[8] As a robustness check we include year-month fixed effects and our yet-to-be presented results remain consistent.
[9] To check the robustness of our findings, we add state-specific linear time trends, and find no substantive change to our corresponding estimates from equation (1).

our standard errors at the county level as well as at the more plausibly stringent state level.[10] After examining our entire sample, we explore the heterogeneity in responses between different types of physicians by observing age, gender, and international medical training. We further subsample by patient characteristics, such as education, and race.

As is well-known, the validity of difference-in-differences estimates relies on the strong assumption that, before the TBT reversal, the C-sections in both groups evolved the same way and, after the TBT reversal, would have evolved the same way if the TBT was not overturned. Appendix Figure B1 provides visual evidence that the parallel paths assumption is maintained for the levels of C-section delivery prior to the TBT reversal. Looking more closely at the dynamics for low Apgar score, and low birth weight using event-study approaches (see Appendix B), we, however, cannot conclude any systematic or lasting impacts on infant health. As shown in Appendix Figures B2 and B3, the results for low Apgar and low birth weight must be interpreted in a rather conservative way, since it is not clear that the parallel paths assumption hold as well as it does for C-sections.

3. Results:

Table 2 displays the effect of the TBT results on C-sections, low Apgar scores and low birth weight using equation (1) for all four samples (Groups A-D). The overall results illustrate that the initial TBT findings published in late 2000 had no statistically significant effect on C-sections (panel A). This is perhaps unsurprising given the research confirmed pre-existing beliefs on the use of C-section for breech births. Similar to us, Jensen and Wüst (2015) found the initial TBT had no effect on C-sections for first-time mothers and suggested this may have been due to first-time breech birth mothers already having a high C-section rate (above 80 percent). Perhaps

[10] Estimates which are statistically significant when clustered at the county level remain precisely estimated when we cluster at the state level, hence we report the latter in our tables given that we use state level variables in our model.

the same logic follows for our findings since the C-section rate was already high for breech births prior to the initial TBT results. Appendix Figure B1 clearly shows C-section rates for breech singleton babies ranged between 80-85 percent prior to the initial TBT study. However, our findings for the TBT reversal are very different: for all definitions of our treatment group, a large and statistically significant reduction in C-section rates is estimated. Looking across the columns in Panel B, it can be seen that, removing factors which increase the likelihood of a C-section leads to a larger decline in the use of C-sections for breech deliveries. This suggests that the marginal birth affected by the new evidence conveyed after 2004 mostly affected less complicated breech births, on which physicians probably have greater discretion in terms of the method of delivery. In our preferred sample, Group D, the reversed TBT results drop the use of C-sections on breech deliveries by 23 percent.[11,12] We note that those results remain whether we exclude the year 2000-2004 or not from the "pre" period in our analysis.

The last column in Table 2 displays the results from a falsification test. We construct a falsification test comprised of "risky" pregnancies. Any evidence suggesting an association between the initial TBT results and its reversal for this category of births, which should not be impacted in an important way given that they are often cases where C-sections are medically indicated, would raise doubts as to the causal inference that we can draw from results in our main sample. In this test, removing breech births completely, the women categorized as having a risky pregnancy were considered the treatment group and the women categorized as non-risky were the control group. Our falsification test shows statistically significant coefficient signs all in the opposite direction and speaks to the increasing trend in C-sections for risky pregnancies, as

[11] Appendix D displays a progression of equation (1) for Group D. This table illustrates that the results are not especially sensitive to the addition of controls, as well as state-specific linear time trends, which we include as a robustness check.
[12] To confirm that our main results are not driven by outliers, we perform a sensitivity analysis by removing high and low C-section rate states and counties. Appendix E shows results do not substantially vary from our main estimates.

opposed to our breech deliveries which display negative coefficients. We note that the results for this falsification test are approximately ten times smaller, and a careful look at Appendix C show no important change for risky pregnancies after 2004.

Panels C and D of Table 2 report the regression results from equation (1) estimated with, respectively, low Apgar scores and low birth weight as outcomes variables. Our difference-in-differences estimates point to a small but statistically significant incidence of low Apgar scores. Starting in 2004, we also document a modest reduction in cases of low birth weight. These muted associations are consistent with many findings in the medical malpractice literature which suggest that changes in the use of C-sections do not coincide with clear changes in infant health (see Currie and MacLeod, 2008; Frakes, 2012; Malak and Yang, 2019; Sloan et al., 1995; Yang et al., 2012). Perhaps because of this lack of precision in the data, we observe more muted, and opposite, effects on infant health based on these two outcomes. First, panel C suggests that the reversal of the TBT led to a increase of a little more than 2 percentage points in the probability of a low Apgar score for term singleton breech newborns. This could happen if the marginal breech birth delivered vaginally after 2004 were not the ideal candidates. For example, it might be the case that physicians did not acquire sufficient information on how to select good candidates for vaginal births, especially since expertise for vaginal breech births was already low in the US (13% at the time of the Trial according to Kotaska (2004)). Comparing columns across panel C, we also note that the increase in low Apgar scores seems to be more relatively homogenous across riskier breech births and across parity statuses, although the results are only statistically significant at the 5% level for groups A and B. The event study graphs in Appendix Figure B5 however fail to show a statistically significant effect for any of our treatment groups. We therefore interpret this result with caution.

Looking at another measure of infant health, low birth weight, Panel D of Table 2 provides a somewhat different picture. The reversal of the TBT is associated with a reduction between five and thirteen percent of breech newborns weighing less than 2500 grams between 2005 and 2010 (a 0.2 to 0.5 percentage point change). As with C-sections, the change is more important in terms of magnitude when focusing on the more restricted sample (D). Overall, this outcome could be a mechanical consequence of the results displayed in Panel B. As more patients deliver vaginally rather than by elective C-section, the gestation may be slightly longer, allowing the baby to grow in the womb. In this case again, however, event study graphs show no statistically significant impact, and we therefore do not put too much stock in the small effects we observe in our difference-in-differences model.

We present estimates from our models based on physician's gender, age, and international medical training in Table 3.[13] As can be seen in the first two columns, there is little difference in the percentage change in C-section use across the genders. The greatest difference estimated is between young and old physicians where we see almost double the drop in C-section use in counties with younger physicians, which implies that it is the younger cohort that is effectively updating their beliefs and closely following medical research updates. Furthermore, it seems that counties with more physicians from international medical schools also decreased their use of C-sections more than counties with obstetricians that only attended U.S. medical schools. Finally, in Table 4 we also look at heterogeneous effects by patient's education and race and find a larger impact on non-white, and less educated patients. Appendix H provides an in-depth discussion of these results.

---

[13] Using the 2005 American Medical Association Physician Professional Data, we calculate county averages for obstetricians and obstetrician-gynecologists. We focus on these characteristics since prior work suggests that practice patterns vary across them (c.f., Kaiser, 2017; Tsugawa et al., 2017; and Woodward et al., 1996).

4. Conclusion:

When the TBT results were reversed and the conclusion that all breech deliveries need not be C-sections was later disseminated among the medical community, we find an arguably large reduction in C-sections for breech births. This strong response to the overturned TBT results also came at a time when the overall trend in C-sections was increasing. Contrary to prior studies, we find that physicians updated their beliefs quickly, and do indeed adjust to new medical research, particularly young physicians. Thus illustrating particular physicians respond to research evidence prior to mandatory policy or professional guidelines.

Doyle et al. (2010) suggest that variation in physician procedural choice may arise because some providers are more competent than others. Breech vaginal deliveries specifically fall in this category, where manoeuvres (ie. external cephalic version) required for vaginal term breech deliveries are seldom taught to residents. The rate of C-section delivery for breech births was only 14 percent in 1970 and had steadily increased up to 85 percent. According to Coco and Silverman (1998) following certain selection criteria for which breech births should attempt a vaginal delivery would lead to about a 60 percent C-section rate for breech births. Another study found that using this manoeuvre to rotate the breech infant would result in 12.3 percent cost savings per birth, in addition to the mother not needing to go through a C-section surgery (Zhang et al., 1993). It is clear that C-section rates for breech births had soared and were clearly above the optimal rate. Many physicians have argued that the extinction of trained obstetricians in vaginal breech delivery is occurring. Our research shows that physicians are paying attention to medical research even when it is counter to strong pre-existing beliefs.

Work Cited

Alarab, M., Regan, C., O'Connell, M.P. *et al* . "Singleton vaginal breech delivery at term: Still a safe option." *Obstet Gynecol* 2004; 103: 407– 412.

Alexander, D. 2015. "Does Physician Compensation Affect Procedure Choice and PatientHealth? Evidence from Medicaid C-section Use." Center for Health and Wellbeing Working Paper. Available at http://scholar.princeton.edu/dalexand/data-1.

Alexandersson, O., Bixo, M., Hogberg, U. 2005. "Evidence-based changes in term breech delivery practice in Sweden."*Acta Obstetricia et Gynecologica Scandinavica* 84: 584–587.

Almond, D., Currie, J., Simeonova, E. 2011. "Public vs. Private Provision of Charity Care? Evidence of the Expiration of Hill-Burton Requirements in Florida." *Journal of Health Economics*, 30(1), 189-199.

Almond, D., and Doyle, J. 2011. "After Midnight: A Regression Discontinuity Design in the Length of Postpartum Hospital Stays." *American Economic Journal: Economic Policy*, 3(3), 1-34.

Almond, D., Doyle, J., Kowalski, A., Williams, H. 2010. "Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns." *Quarterly Journal of Economics*, 125(2), 591-634.

American College of Obstetricians and Gynecologists. 2001. "ACOG committee opinion: no. 265: Mode of term single breech delivery." *Obstetrics and Gynecology* 98(6): 1189 90.

American College of Obstetricians and Gynecologists. 2004. "ACOG Practice bulletin no. 54: Vaginal birth after previous cesarean." *Obstetrics and Gynecology* 104(1): 203-12.

Avraham, R. 2007. "An Empirical Study of the Impact of Tort Reforms on Medical Malpractice Settlement Payments." *Journal of Legal Studies* 36(S2): S183-S229.

Avraham, R. 2014. "Database of State Tort Law Reforms 5th Ed." University of Texas School of Law: Law and Economics Research Paper No.e555.

Baker, M., Isabelle, M., Stabile, M. Allin, S. 2022. "Accounting for the Rising Caesarean Section Rate in Canada: What Are the Roles of Changing Needs, Practices, and Incentives?" *Canadian Public Policy* 48(1):36-73.

Berez, J., David, G., Howard, D. H., Neuman, M. D. 2018. "Does bad news travel faster? On the determinants of medical technology abandonment." *Journal of Human Capital* 12(4):569-603.

Bertoli, P., Grembi, V. 2019. "Malpractice risk and medical treatment selection." *Journal of Public Economics* 174:22-35.

Betrán, A.P., Ye, J., Moller, A-B., Zhang, J., Gülmezoglu, A.M., Torloni, M.R. 2016."The Increasing Trend in Caesarean Section Rates: Global, Regional and National Estimates: 1990-2014." *PLoS ONE* 11(2): e0148343.

Brown, H S. 1996. "Physician Demand for Leisure: Implications for Cesarean Section Rates." *Journal of Health Economics* 15(2); 233-242.

Burstin, H.R., Johnson, W.G., Lipsitz, S.R., Brennan, T.A. 1993. "Do the poor sue more? A case-control study of malpractice claims and socioeconomic status." *Journal of the American Medical Association* 270:1697–1701.

Coco, A.S., Silverman, S.D. 1998. "External Cephalic Version." *American Family Physician*, 58(3):731-738.

Currie, J., Gruber, J., Fischer, M. 1995. "Physician Payments and Infant Mortality: Evidence from Medicaid Fee Policy." *American Economic Review*, 85(2): 106-111.

Currie, J., Lin, W., Zhang, W. 2010. "Patient Knowledge and Antibiotic Abuse: Evidence from an Audit Study in China." *Journal of Health Economics*, 30(5): 933-949.

Currie, J., MacLeod, W. B. 2008. "First do no harm? Tort reform and birth outcomes." *Quarterly Journal of Economics* 123(2): 795–830.

Currie, J., MacLeod, W. B. 2017. "Diagnosing expertise: Human capital, decision making, and performance among physicians." *Journal of Labor Economics* 35(1): 1-43.

Currie, J. 2011. "Inequality at birth: some causes and consequences." *American Economic Review*: *Papers and Proceedings* 101(3):1-22.

Cutler, D., Skinner, J., Stern, A.D., Wennberg, D. 2019. "Physician beliefs and patient preferences: A new look at regional variation in health care spending." *American Economic Journal: Economic Policy*, 11(1):192-221.

Dehejia, R., Lleras-Muney, A. 2004. "Booms, busts, and babies' health." *Quarterly Journal of Economics* 119: 1091-1130.

Doyle, J.J., Ewer, S.M., Wagner, T.H. 2010. "Returns to provider human capital: Evidence from patients randomized to provider teams." *Journal of Health Economics*, 29, 866-882.

Dubay, L., Kaestner, R., Waidmann, T. 1999. "The Impact of Malpractice Fears on Cesarean Section Rates." *Journal of Health Economics* 18: 491-522.

Epstein, A.J., Nicholson, S. 2009. "The formation and evolution of physician treatment styles: an application to cesarean sections."*Journal of Health Economics* 28 (6): 1126–1140.

Esposto, A. G. 2012. "Tort reform and caesarean deliveries."*Applied Economics Letters* 19(12): 1171–1174.

Fabbri, D., Monfardini, C., Castaldini, I., Protonotari, A. 2016. "Cesarean section and the manipulation of exact delivery time." *Health Policy* 120(7): 780-789.

Frakes, M. 2012. "Defensive medicine and obstetric practices." *Journal of Empirical Legal Studies* 9 (3): 457–481.

Glezerman, M. 2006. "Five years to the term breech trial: the rise and fall of a randomized controlled trial."*American Journal of Obstetrics and Gynecology* 194 (1): 20–25.

Green, A.R., Carney, D.R., Pallin, D.J., Ngo, L.H., Raymond, K.L., Iezzoni, L.I., Banaji, M.R. 2007. "Implicit Bias among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients." *Journal of General Internal Medicine* 22: 1231-1238.

Gruber, J., Kim, J., Mayzlin, D. 1999. "Physician Fees and Procedure Intensity: The Case of Cesarean Sections." *Journal of Health Economics*, 18(4): 473-490.

Goffinet, F., Carayol, M., Foidart, J.M., Alexander, S., Uzan, S., Subtil, D., Breart, G. 2006. "Is planned vaginal delivery for breech presentation at term still an option? Results of an observational prospective survey in France and Belgium." *American Journal of Obstetrics and Gynecology* 194(4): 1002–1011.

Hannah, M.E., Hannah, W.J., Hewson, S.A., Hodnett, E.D., Saiga, S.l., Willan, A.R. 2000. "Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial." *The Lancet* 356 (9239): 1375–1383.

Howard, D. H., David, G., Hockenberry, J. 2017. "Selective hearing: Physician-ownership and physicians' response to new evidence." *Journal of Economics & Management Strategy* 26(1): 152-168.

Jensen, V.M., Wüst, M. 2015. "Can Caesarean section improve child and maternal health? The case of breech babies." *Journal of Health Economics* 39: 289–302.

Johnson, E. M., and M. M. Rehavi. 2016. "Physicians Treating Physicians: Information and Incentives in Childbirth." *American Economic Journal: Economic Policy* 8 (1): 115-41.

Kaiser, B. 2017. "Gender-specific practice styles and ambulatory health care expenditures." *European Journal of Health Economics* 18: 1157–1179.

Kessler, D., McClellan, M. 2002. "Malpractice law and health care reform: optimal liability policy in an era of managed care" *Journal of Public Economics*, 84: 175-197.

Kotaska, A. 2004. "Inappropriate use of randomised trials to evaluate complex phenomena: case study of vaginal breech delivery." *British Medical Journal* 329 (7473): 1039–1042.

Malak, N., Yang, Y. T. 2019. "A re-examination of the effects of tort reforms on obstetrical procedures and health outcomes." *Economics Letters* 184, 108626.

McClellan, F. M., White, A. A., Jimenez, R. L., Fahmy, S. 2012. "Do Poor People Sue Doctors More Frequently? Confronting Unconscious Bias and the Role of Cultural Competency." *Clinical Orthopaedics and Related Research* 470(5): 1393–1397.

Moster, Dag, Lie, Rolv T., Irgens, Lorentz M., Tor Bjerkedal, Trond Markestad. 2001. "The association of Apgar score with subsequent death and cerebral palsy: a population-based study in term infants."*Journal of Pediatrics* 138 (6), 798–803.

Nelson, K.B., Ellenberg, J.H. 1981. "Apgar scores as predictors of chronic neurologic disability." *Pediatrics* 68 (1), 36–44.

Newhouse, J. P., Garber, A. M. 2013. "Geographic Variation in Medicare Services." *New England Journal of Medicine* 368: 1465-1468.

Oster, E.F. 2018. Expert Behavior Change in Response to Best Practice Changes: Evidence from Obstetrics, Working paper.

Palmer, W., Bottle, A., Aylin, P. 2015. "Association between day of delivery and obstetric outcomes: observational study."*British Medical Journal* 351: h5774.

Phelps, Charles E. 1992. "Diffusion of Information in Medical Care." *Journal of Economic Perspectives*, 6 (3): 23-42.

Price, J., Simon, K. 2009. "Patient education and the impact of new medical research." *Journal of Health Economics* 28 (6): 1166–1174.

Rietberg, C., Elferink-Stinkens P., Visser, G. 2005. "The effect of the Term Breech Trial on medical intervention behaviour and neonatal outcome in The Netherlands: An analysis of 35 453 term breech infants."*BJOG: An International Journal of Obstetrics & Gynaecology* 112:205–209.

Roberts, C.L., Algert, C.S., Peat, B., Henderson-Smart, D. 1999. "Small fetal size: a risk factor for breech birth at term." *International Journal of Gynecology and Obstetrics* 67(1):1-8.

Shurtz, I. 2013. "The impact of medical errors on physician behavior: evidence from malpractice Litigation." *Journal of Health Economics* 32: 331–340.

Sloan, F., Entman, S. S., Reilly, B. A., Glass, C. A., Hickson, G. B., Zhang, H. H. 1995. "Tort liability and obstetricians' care levels." Duke University, Department of Economics, Working Papers: 95-107.

Staats, B. R., Kc, D. S., Gino, F. 2018. "Maintaining beliefs in the face of negative news: The moderating role of experience." *Management Science* 64(2): 804-824.

Sullivan, E. A., Moran, K., Chapman, M. 2009. “Term Breech Singletons and Caesarean Section: A Population Study, Australia 1991-2005”*Australian and New Zealand Journal of Obstetrics and Gynaecology* 49(5):456-460.

Tsugawa, Y., Newhouse, J. P., Zaslavsky, A. M., Blumenthal, D. M., Jena, A. B. 2017. “Physician age and outcomes in elderly patients in hospital in the US: observational study.” *BMJ* 357: j1797.

Turner, M. J. 2006. “The term breech trial: are the clinical guidelines justified by the evidence.” *Journal of Obstetrics and Gynaecology* 26 (6): 491–494.

Woodward, C., Cohen, M., Ferrier, B., Williams, P. 1996. “The Relative Importance of Undergraduate and Postgraduate Education to the Practice Decisions and Attitudes of Young Family Physicians Practising in Ontario.” Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Canada, Centre for Health Economics and Policy Analysis Working Paper Series: 1996-01.

Wu, B., David, G. 2022. “Information, relative skill, and technology abandonment.” *Journal of Health Economics* 83, 102596.

Yang Y. T., Studdert, D. S., Subramanian, S. V., Mello, M. M. 2012. “Does Tort Law Improve the Health of Newborns, or Miscarry? A Longitudinal Analysis of the Effect of Liability Pressure on Birth Outcomes.”*Journal of Empirical Legal Studies* 9(2):217–245.

Yang, Y. T., Mello, M. M., Subramanian, S. V., Studdert, D. M. 2009. “Relationship between malpractice litigation pressure and rates of cesarean section and vaginal birth after caesarean section.” *Medical Care* 47 (2): 234–242.

Zhang, J., Bowes, W.A., Fortney, J.A. 1993. “Efficacy of external cephalic version: a review.” *Obstet Gynecol*, 82:306-312.

Table 1: Summary Statistics for Preferred Sample (Group D)

	Breech Births (Treatment)			Vertex/Head Down Births (Control)		
	Pre-TBT Results (Jan.95-Oct.00)	Initial TBT results (Nov.00-Oct.04)	Reversed TBT results (Nov. 04-Dec.10)	Pre-TBT Results (Jan.95-Oct.00)	Initial TBT results (Nov.00-Oct.04)	Reversed TBT results (Nov. 04-Dec.10)
Outcome Variables:						
C-section	0.784	0.800	0.659	0.066	0.085	0.114
Low 5-Minute Apgar (<9)	0.109	0.112	0.160	0.065	0.067	0.090
Low Birth Weight (<2500g)	0.042	0.039	0.037	0.022	0.021	0.022
Mother's Characteristics:						
White	0.690	0.659	0.540	0.603	0.573	0.544
Hispanic	0.152	0.180	0.287	0.189	0.224	0.248
Black	0.096	0.099	0.111	0.146	0.140	0.139
Teenager	0.044	0.034	0.037	0.059	0.048	0.044
Forty and older	0.035	0.041	0.041	0.023	0.027	0.028
No High School Diploma	0.176	0.178	0.215	0.211	0.214	0.207
Some College and Higher	0.241	0.285	0.305	0.221	0.250	0.291
Married	0.751	0.736	0.661	0.716	0.696	0.644
Pregnancy Characteristics:						
Smoked during Pregnancy	0.168	0.139	0.107	0.142	0.122	0.108
Newborn Boy	0.477	0.479	0.488	0.508	0.509	0.509
Number of Prenatal Visits	11.646	11.544	11.242	11.471	11.444	11.116
State Characteristics:						
Medicaid Fee Difference	69.721	79.126	69.622	71.666	82.474	85.538
County Unemployment Rate	4.864	5.473	6.488	5.013	5.566	6.414
Caps on NED	0.344	0.361	0.637	0.349	0.367	0.571
Caps on PD	0.500	0.546	0.643	0.499	0.548	0.616
JSL reform	0.726	0.761	0.828	0.738	0.762	0.787
CSR	0.621	0.682	0.642	0.630	0.665	0.697
Number of Observations	233536	161640	356944	10182836	6858058	10267163

Note: The sample period is from January 1995 to December 2010. This is for Group D which consists of non-risky term singleton births with no previous C-sections and first parity births removed. The initial Term Breech Trial (TBT) results were published in October 2000, the reversed TBT results were published in October 2004. NED refers to non-economic damages, PD refers to punitive damages, JSL refers to the joint and several liability rule, and CSR refers to the collateral source rule.

Table 2: Effects of TBT on C-Sections and Infant Health Outcomes

	Group A	Group B	Group C	Group D	Falsification Test (Risky Pregnancy)
Panel A: C-section Outcome During Initial TBT Results					
Coefficient β_1	-0.011	-0.010	-0.005	-0.006	0.011
Standard Error	0.007	0.007	0.009	0.009	0.001
P-value	0.095	0.148	0.563	0.498	0.000
Percentage Change (%)	-1.3	-1.3	-0.7	-0.8	1.4
Observations	10,366,741	10,290,752	10,066,080	10,044,080	10,663,106
Panel B: C-section Outcome During Reversed TBT Results					
Coefficient β_1	-0.124	-0.149	-0.172	-0.182	0.016
Standard Error	0.051	0.058	0.070	0.072	0.002
P-value	0.020	0.014	0.019	0.016	0.000
Percentage Change (%)	-14.6	-18.0	-21.9	-23.2	9.3
Observations	13,357,137	13,209,465	12,847,666	12,810,002	13,711,803
Panel C: Low 5-Minute Apgar (<9) Outcome During Reversed TBT Results					
Coefficient β_1	0.019	0.020	0.018	0.021	-0.009
Standard Error	0.009	0.009	0.010	0.010	0.002
P-value	0.047	0.033	0.078	0.054	0.000
Percentage Change (%)	18.0	19.5	15.7	19.9	-8.9
Observations	12,399,699	12,259,035	11,915,900	11,879,983	12,728,343
Panel D: Low Birth Weight (<2500g) Outcome During Reversed TBT Results					
Coefficient β_1	-0.003	-0.002	-0.005	-0.005	-0.004
Standard Error	0.002	0.002	0.002	0.002	0.001
P-value	0.091	0.129	0.021	0.008	0.000
Percentage Change (%)	-6.4	-5.6	-10.9	-12.8	-8.1
Observations	13,357,137	13,209,465	12,847,666	12,810,002	13,711,803
Panel E: Group Definition					
Term Singleton	X	X	X	X	X
No Previous C-section		X	X	X	X
Parity >1			X	X	X
Non-Risky Pregnancy				X	

Note: The initial Term Breech Trial (TBT) results were published in October 2000, the reversed TBT results were published in October 2004. Pre-TBT: Jan.95-Oct.00; Initial TBT: Nov.00-Oct.04; Reversed TBT: Nov.04-Dec.10. Panel A does not include the reversed TBT period, and Panels B, C, and D do not include the initial TBT period, although its inclusion does not change estimates. Each panel and column is from a separate regression. The linear probability model includes day of the week, month, year, and county fixed-effects, demographic, pregnancy, and state controls listed in Table 1. Standard errors are clustered at the state level. The percentage change is the coefficient divided by the pre-TBT (prior to October 2000) C-section rate for breech deliveries.

Table 3: Effects of Reversed TBT on C-Sections and Infant Health Outcomes Subsampled by Obstetricians' Characteristics

	Counties with male OBs (1)	Counties with more female OBs (2)	Counties with younger OBs (3)	Counties with older OBs (4)	Counties with more OBs from outside US medical schools (5)	Counties with OBs from US medical schools (6)
Panel A: C-section Outcome During Reversed TBT Results						
Coefficient β_1	-0.119	-0.124	-0.153	-0.075	-0.197	-0.135
Standard Error	0.047	0.045	0.055	0.042	0.086	0.057
P-value	0.018	0.009	0.008	0.083	0.030	0.025
Percentage Change (%)	-15.3	-16.3	-20.0	-9.6	-24.2	-17.4
Observations	714587	3179942	4325513	1033539	3046014	1973254
Panel B: Low 5-Minute Apgar (<9) Outcome During Reversed TBT Results						
Coefficient β_1	0.016	0.022	0.023	-0.003	-0.010	0.024
Standard Error	0.007	0.018	0.011	0.006	0.006	0.013
P-value	0.031	0.222	0.037	0.635	0.122	0.070
Percentage Change (%)	13.3	20.0	20.0	-2.6	-11.2	19.9
Observations	670630	3105945	3982676	993663	2917693	1862541
Panel C: Low Birth Weight (<2500g) Outcome During Reversed TBT Results						
Coefficient β_1	-0.004	-0.003	-0.004	-0.004	-0.006	-0.005
Standard Error	0.005	0.002	0.002	0.003	0.003	0.002
P-value	0.368	0.299	0.075	0.137	0.094	0.037
Percentage Change (%)	-9.2	-6.1	-8.0	-9.4	-14.7	-11.0
Observations	714587	3179942	4325513	1033539	3046014	1973254

Note: This is for Group D which consists of non-risky term singleton births with no previous C-sections and first parity births removed. The columns of different county averages comprise of the top and bottom 25th percentile of differing obstetricians' characteristics in 2005. Column (1) consists of counties with only male obstetricians; column (2) represents counties with over fifty percent female obstetricians. Column (3) represents counties where the average age of obstetricians is less than 46, whereas Column (4) consists of counties where the average age is over 52. Column (5) represents counties where more than twenty-five percent of the obstetricians attended medical school outside of the U.S., and Column (6) represents counties where all the obstetricians were trained in U.S. medical schools. The reversed Term Breech Trial (TBT) results were published in October 2004. Pre-TBT: Jan.95-Oct.00; Initial TBT: Nov.00-Oct.04; Reversed TBT: Nov.04-Dec.10. Panel A, B, and C do not include the initial TBT period, although its inclusion does not change estimates. Each panel and column is from a separate regression. The linear probability model includes day of the week, month, year, and county fixed-effects, demographic, pregnancy, and state controls listed in Table 1. Standard errors are clustered at the state level. The percentage change is the coefficient divided by the pre-TBT (prior to October 2000) C-section rate for breech deliveries.

Table 4: Effects of Reversed TBT on C-Sections and Infant Health Outcomes Subsampled by Mother's Education and Race

	No High School Diploma	Some College and Higher	White	Black	Hispanic
Panel A: C-section Outcome During Reversed TBT Results					
Coefficient β_1	-0.257	-0.124	-0.117	-0.231	-0.307
Standard Error	0.087	0.056	0.044	0.064	0.099
P-value	0.006	0.033	0.013	0.001	0.004
Percentage Change (%)	-33.0	-15.8	-15.0	-29.7	-37.4
Observations	2653731	3375762	7529681	1733254	2795457
Panel B: Low 5-Minute Apgar (<9) Outcome During Reversed TBT Results					
Coefficient β_1	0.0004	0.033	0.024	0.008	0.001
Standard Error	0.011	0.017	0.012	0.015	0.014
P-value	0.973	0.058	0.056	0.620	0.922
Percentage Change (%)	0.3	36.7	23.1	5.3	1.3
Observations	2354035	3208346	7134348	1632377	2394079
Panel C: Low Birth Weight (<2500g) Outcome During Reversed TBT Results					
Coefficient β_1	-0.014	-0.001	-0.003	-0.012	-0.010
Standard Error	0.003	0.001	0.002	0.004	0.001
P-value	0.000	0.374	0.080	0.002	0.000
Percentage Change (%)	-22.3	-4.6	-7.6	-15.2	-25.4
Observations	2653731	3375762	7529681	1733254	2795457

Note: This is for Group D which consists of non-risky term singleton births with no previous C-sections and first parity births removed. The reversed Term Breech Trial (TBT) results were published in October 2004. Pre-TBT: Jan.95-Oct.00; Initial TBT: Nov.00-Oct.04; Reversed TBT: Nov.04-Dec.10. Panel A, B, and C do not include the initial TBT period, although its inclusion does not change estimates. Each panel and column is from a separate regression. The linear probability model includes day of the week, month, year, and county fixed-effects, demographic, pregnancy, and state controls listed in Table 1. Standard errors are clustered at the state level. The percentage change is the coefficient divided by the pre-TBT (prior to October 2000) C-section rate for breech deliveries.

Appendix A: Data Description

Natality Data:

Our primary data come from the United States Birth Certificate Records from 1995-2010 for all states and the District of Columbia. Since the TBT focused exclusively on full-term singleton births, we exclude multiple births and premature births (i.e., babies born before 37 weeks of gestation), which respectively represent three and twelve percent of all births over the period covered by our analysis.

Our main outcome variable is whether or not a mother gave birth via C-section. In addition to modeling the probability of a C-section we also investigate whether there is any detrimental impact on the health of the newborn by examining the incidence of low Apgar scores and low birth weight. An Apgar score is a measure from zero to ten that quickly summarizes the health of a baby five minutes after it is born. It stands for Appearance, Pulse, Grimace, Activity, and Respiration. We define a low Apgar score as any value less than 9 out of 10. In line with the literature, our indicator for low birth weight flags newborns weighing less than 2500 grams at delivery.

Though our data contain a limited set of individual and household characteristics, our explanatory variables include standard demographic information such as race, mother's education, marital status, and age indicators. In addition to demographic variables regarding the mother, we also include some characteristics of the pregnancy and of the newborn in some models. For example, though likely endogenous, we can include for the number of prenatal health care visits and whether or not the mother smoked during the pregnancy in some specifications. Other available controls include birth parity, the baby's sex, and the day of the week the baby was born.[14]

Our data also include information on common risk factors prior to or during pregnancy: chronic diabetes, gestational diabetes, chronic hypertension, gestational hypertension, or eclampsia.[15] Since these conditions are known prior to labor, they may affect physicians' procedural choices. In addition to controlling for the riskiness of a woman's pregnancy in our main models, we also conduct a falsification test where "risky" pregnancies, which differ from the marginal cases on which the studies like the TBT and its critics were focused, act as a "placebo group" for breech births.

Additional Data:

In addition to individual level characteristics available in the Natality Detail Files, we also include key aggregate level controls which prior work suggests might be strong predictors of procedural choice and birth outcomes. First, we control for annual county-level unemployment rates from the labor force data collected by the Bureau of Labor Statistics. It is important to control for business cycle fluctuations as several studies have shown that they may affect health outcomes (e.g., a prominent study by Dehejia and Lleras-Muney (2004) find evidence that economic conditions, as proxied by unemployment rates, systematically affect birth outcomes).

Second, physician behavior/procedural choice may be influenced by pricing differences between vaginal and C-section births. Indeed, Alexander (2015) finds that when Medicaid pays

[14] Descriptive research suggests that babies born on weekends experience worse health outcomes (Palmer et al., 2015).

[15] Diabetes is a metabolic disease in which the body's inability to produce any or enough insulin causes elevated levels of glucose in the blood. Chronic hypertension is defined as a long-term condition of having blood pressure greater than 140/90 mm Hg. Pregnancy hypertension is the development of new hypertension in a pregnant woman after 20 weeks of gestation without the presence of protein in the urine or other signs of pre-eclampsia. Eclampsia is a condition in which one or more convulsions occur in a pregnant woman suffering from high blood pressure, often followed by coma and posing a threat to the health of both mother and baby.

physicians relatively more for C-sections, then they are utilized more often. To avoid any systematic differences in the relative price of C-sections before and after the TBT and its reversal to introduce a bias in our analysis, we include a state-year level variable that accounts for the difference in the Medicaid fee schedule for vaginal and C-section deliveries.[16] While not all births in our sample are covered by Medicaid, this variable can act as a proxy for pricing differences in each state over time.

Finally, we use the fifth edition of the *Database of State Tort Law Reforms* to characterize tort reforms related to medical malpractice. This database contains the most detailed, complete, and comprehensive information on state tort reforms.[17] All of our models control for common tort reforms that might affect physicians' decisions regarding whether or not to provide a C-section.[18] In particular, we include state-year level indicator variables for caps on non-economic damages (NED), caps on punitive damage (PD), the elimination of the joint-and-several liability (JSL) rule, and the collateral source rule (CSR) to characterize the liability environment faced by physicians.[19,20]

---

[16] Medicaid fee schedule was provided by Alexander (2015).
[17] This database has become the gold standard in empirical tort reform research by providing a review of original legislation and case law with exact text and effective dates. See Avraham's (2014) Database of State Tort Law Reforms (5th) for more information on the tort reform database.
[18] See Bertoli and Grembi (2019); Currie and MacLeod (2008); Dubay et al. (1999); Esposto (2012); Shurtz (2013); and Yang et al. (2009).
[19] Consistent with prior work in this area, indicators are assigned based on the state and birth year of the infant.
[20] See Malak and Yang (2019) for a description of these popular tort reforms and their effects on maternal and infant health.

Appendix B: Event-Studies

In addition to the regression tables, we also conduct event-studies. We construct figures to show how our main outcomes of interest evolve before and after the overturned TBT results by plotting the a^j coefficients from the following regression:

$$(A1) \qquad Y_{it} = a + \sum_{j=-6}^{-2} a^j \bullet 1\{event_{st} = j\} + \sum_{j=0}^{6} a^j \bullet 1\{event_{st} = j\} + X_{it}^{'}\lambda + \gamma_c + \gamma_t + u_{it}$$

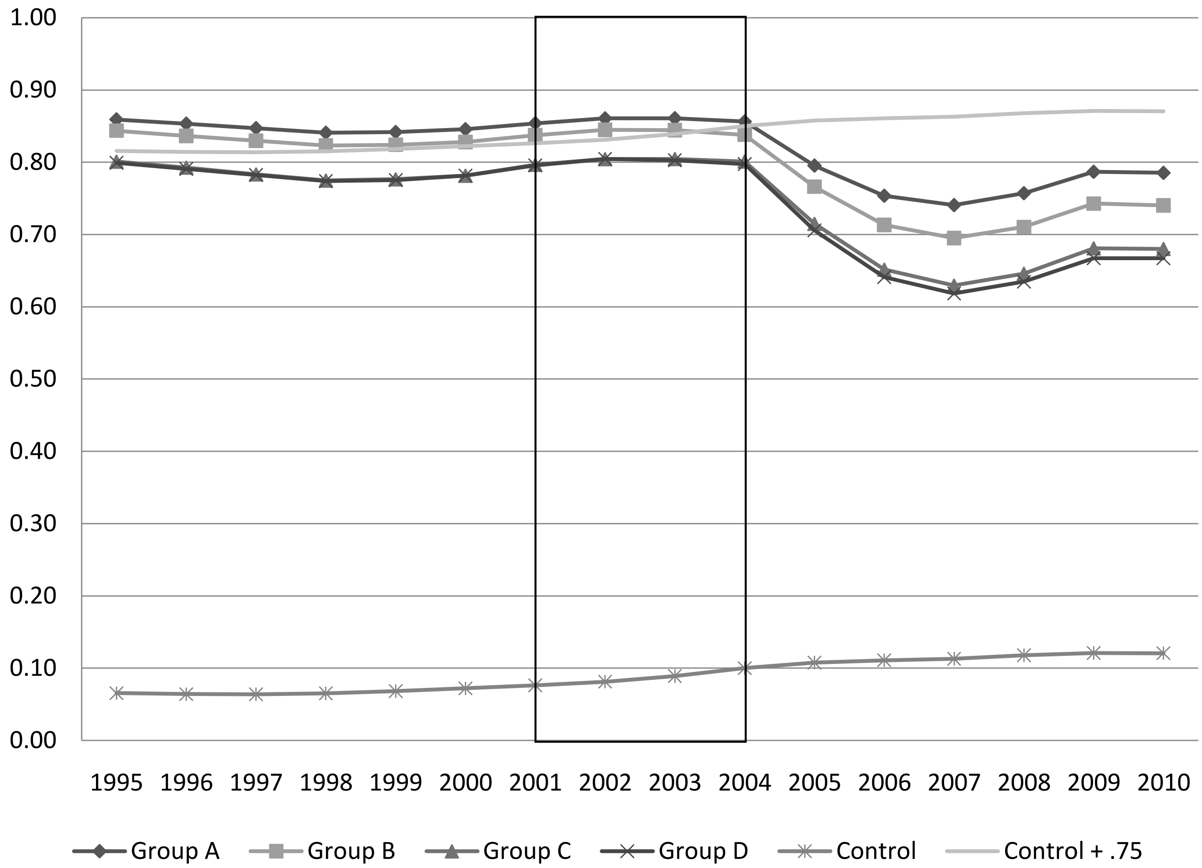
where $1\{event_{st} = j\}$ is an indicator variable equal to one if the treatment state-year is in event time j, and equal to zero otherwise. Time 0 in the event-study is year 2005 since the reversed TBT results were published at the end of 2004. The event time variable stops at ten years pre and five years post due to our data sample (1995-2010). The outcomes of interest, Y_{it}, are the same as equation (1): C-section delivery, low Apgar, and low birth weight for individual i, in year t. This specification mirrors equation (1) where we still keep all the controls in vector $X_{it}^{'}$ for individual i in year t. Also, we keep county fixed effects, γ_c, and year fixed effects, γ_t. Standard errors remain clustered at the state level.

In Appendix Figure B1, we first investigate the parallel paths assumption for the C-section level, which seems to hold reasonably well between 1995 and 2000. The assumption seems to continue to hold between 2000 and 2004, without any apparent break in the trend for neither of the treatment or control groups. This suggests that the publication of the TBT may not have had a significant impact on US physicians' propensity to opt for a C-section when delivering term breech babies, although the evidence from Appendix Figure B1 remains observational and cannot be given a causal interpretation. Since there is a well-documented rise in C-sections over time (Betrán, 2016), we also present in Appendix Figure B4 event-studies for each of the treatment groups we consider, with 2005 as the event (time 0). In this case again, the parallel path assumption seems to hold for the period spanning from 1995 to 2004.

Appendix Figures B2 and B3 present the unconditional levels for our two infant health measures. For low Apgar scores (Figure B2), the parallel path assumption seems to hold for the first period (1995 to 2000). The release of the initial TBT results do not appear to have caused breech births to diverge too much, although a relative increase in the incidence of low Apgar scores for all our treatment groups can be observed between 2003 and 2004. However, this difference is not statistically significant in the event studies (Figure B5).[21] An upward trend is visible for the period following the reversal of the TBT results, between 2004 and 2010, although event-study estimates are noisy. The raw evolution of the proportion of births below 2500 grams is presented in Figure B3. In this case, the parallel path assumption seems less convincing for the period spanning from 1995 to 2000, when low birth weight seems to rise for breech births compared to the control groups. However, rates look relatively stable for all groups between 2000 and 2004, and seem to rise again after 2004, but only for breech births. Noisier visual evidence presented in the corresponding event study graphs (Figure B6) goes in the same direction.

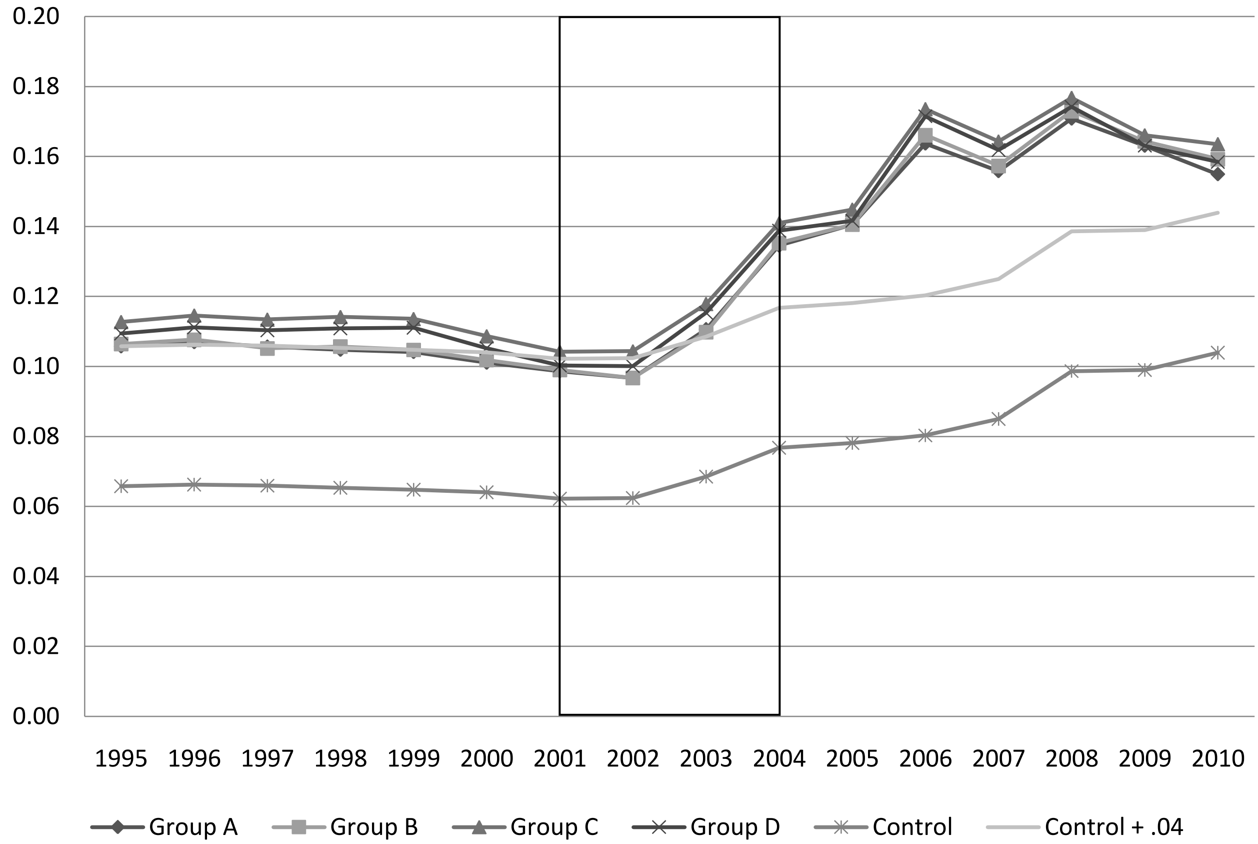
[21] The movement between 2003 and 2004 could also be due to the fact that a few months can be considered to be "treated" in 2004. One of the main studies contesting the TBT conclusions, Kotaska (2004), was published in October of 2004.

Appendix Figure B1: C-Section Rate Between Breech (Treatment) and Vertex/Head Down (Control) Deliveries



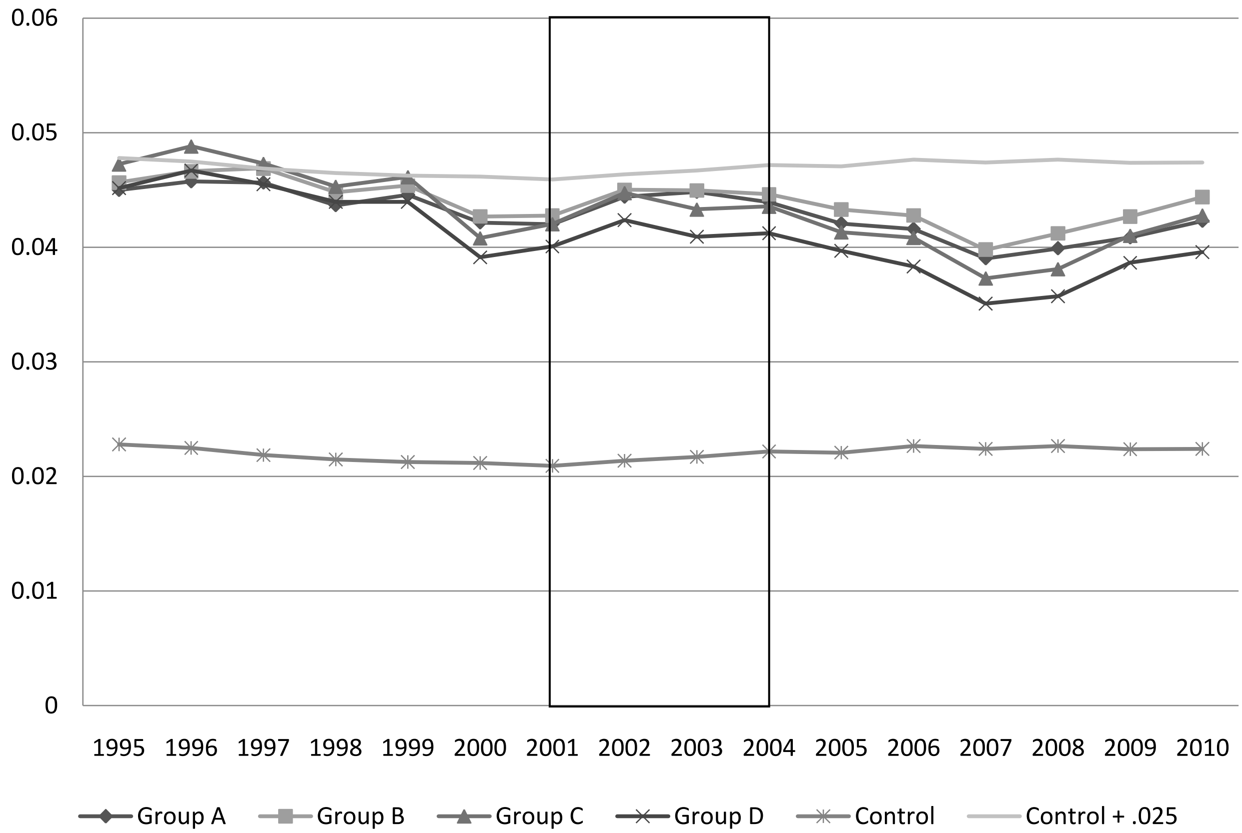
Note: The initial Term Breech Trial (TBT) results were published in October 2000, the reversed TBT results were published in October 2004. Group A are term singleton breech births. Group B builds on Group A and excludes those with previous C-sections. Group C also removes first parity births, and finally, group D also removes those characterized with risky pregnancies (ie. Chronic or gestational diabetes, blood pressure, or eclampsia).

Appendix Figure B2: Low 5-Minute Apgar Scores (<9) Between Breech (Treatment) and Vertex/Head Down (Control) Deliveries



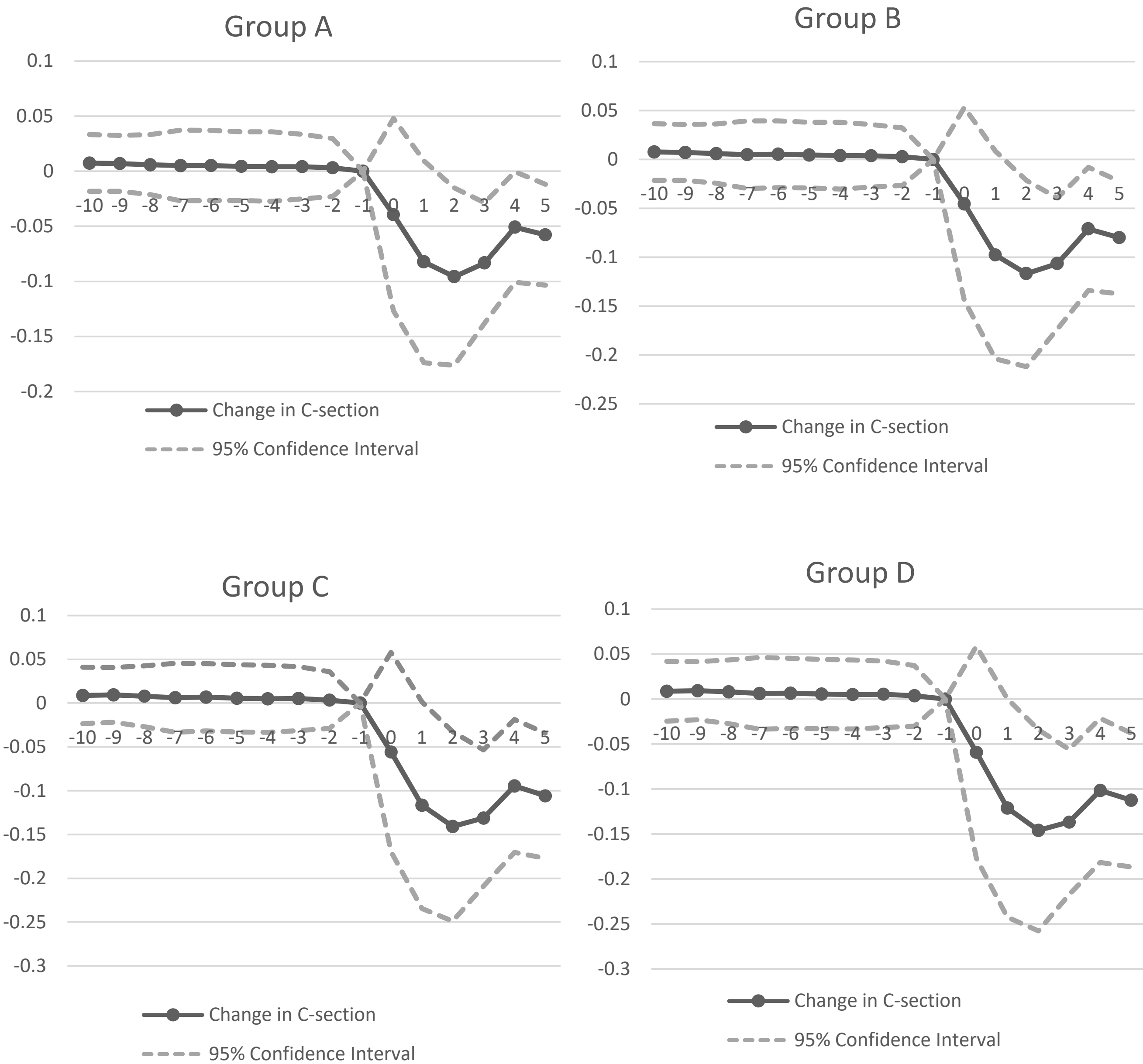
Note: The initial Term Breech Trial (TBT) results were published in October 2000, the reversed TBT results were published in October 2004. Group A are term singleton breech births. Group B builds on Group A and excludes those with previous C-sections. Group C also removes first parity births, and finally, group D also removes those characterized with risky pregnancies (ie. Chronic or gestational diabetes, blood pressure, or eclampsia). In 2003 the standard U.S. birth certificate was revised to now include a 10-minute Apgar score in addition to the 5-minute Apgar score and further removed the 1-minute Apgar score. We believe these changes along with an ACOG Committee Opinion in 2006 to use an expanded Apgar score reporting form can help account for the increase in low Apgar scores over time.

Appendix Figure B3: Low Birth Weight (<2500g) Between Breech (Treatment) and Vertex/Head Down (Control) Deliveries



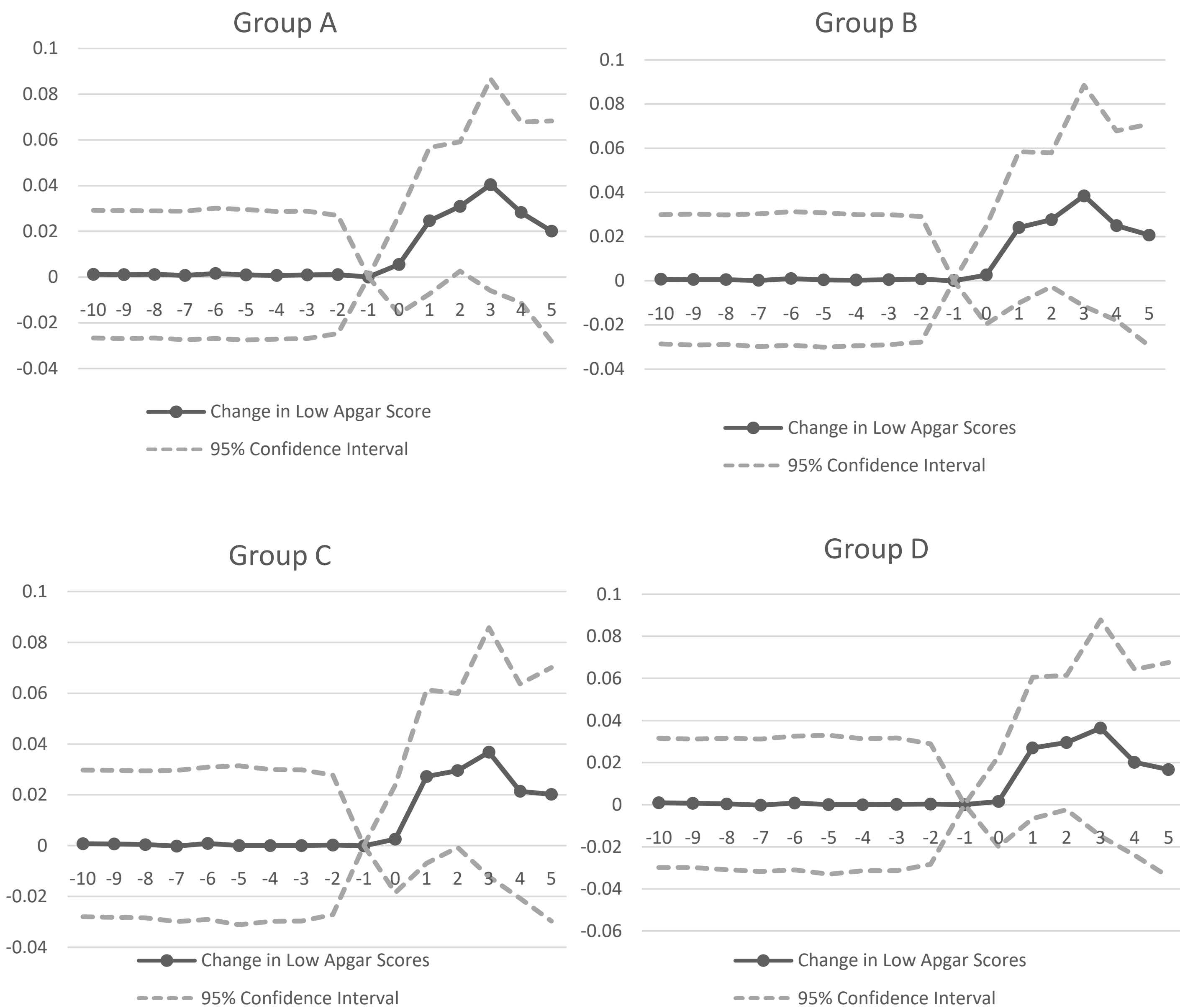
Note: The initial Term Breech Trial (TBT) results were published in October 2000, the reversed TBT results were published in October 2004. Group A are term singleton breech births. Group B builds on Group A and excludes those with previous C-sections. Group C also removes first parity births, and finally, group D also removes those characterized with risky pregnancies (ie. Chronic or gestational diabetes, blood pressure, or eclampsia).

Appendix Figure B4: Change in C-Section- Event-Study for Breech Deliveries



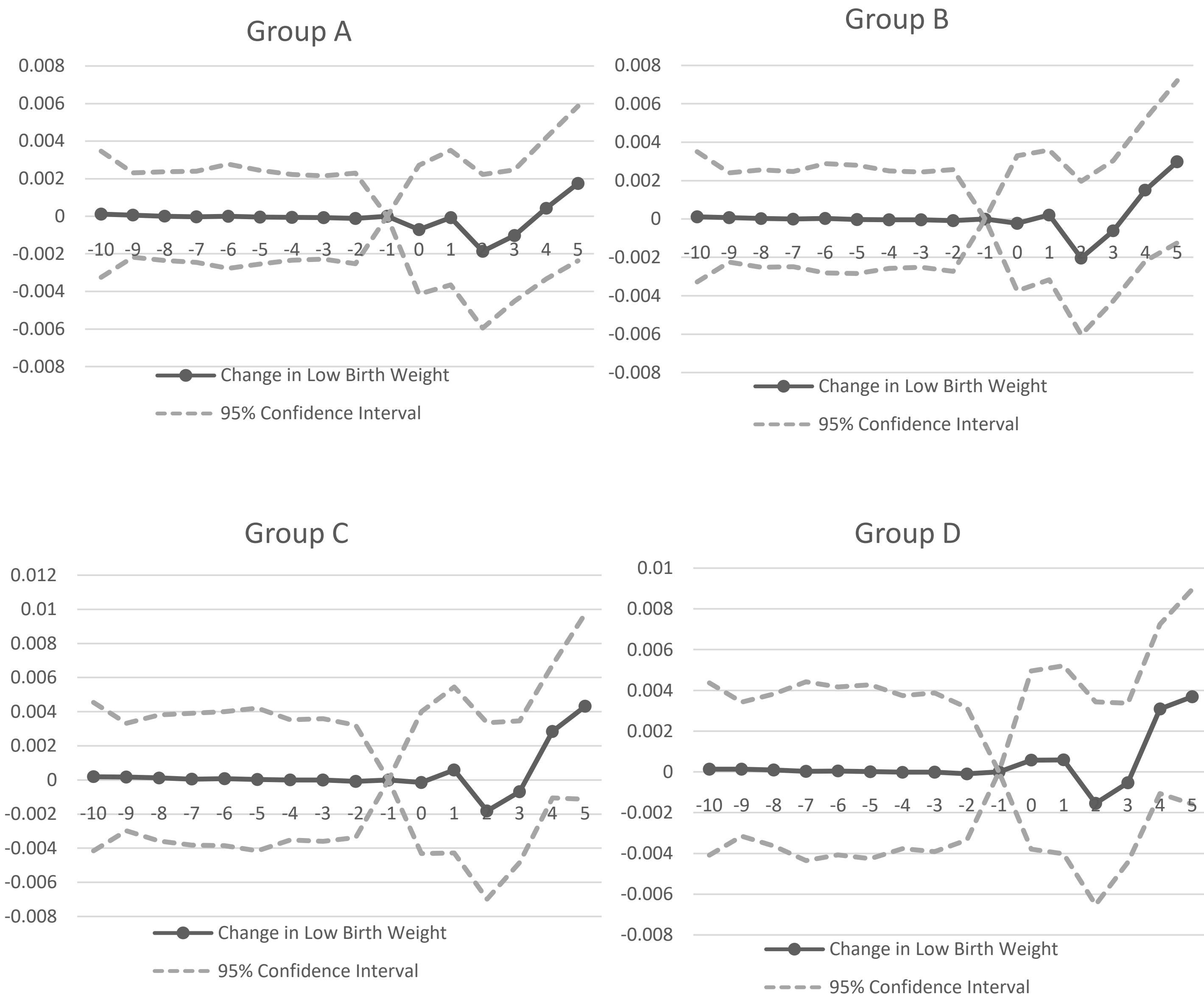
Notes: The initial Term Breech Trial (TBT) results were published in October 2000, the reversed TBT results were published in October 2004. Time 0 in the event-studies is year 2005 since the reversed TBT results were published at the end of 2004. Each dark marker is an estimate of the a^j coefficient from equation (A1) with the confidence interval on either side. Group A are term singleton breech births. Group B builds on Group A and excludes those with previous C-sections. Group C also removes first parity births, and finally, group D also removes those characterized with risky pregnancies (ie. Chronic or gestational diabetes, blood pressure, or eclampsia).

Appendix Figure B5: Change in Low 5-Minute Apgar Scores- Event-Study for Breech Deliveries



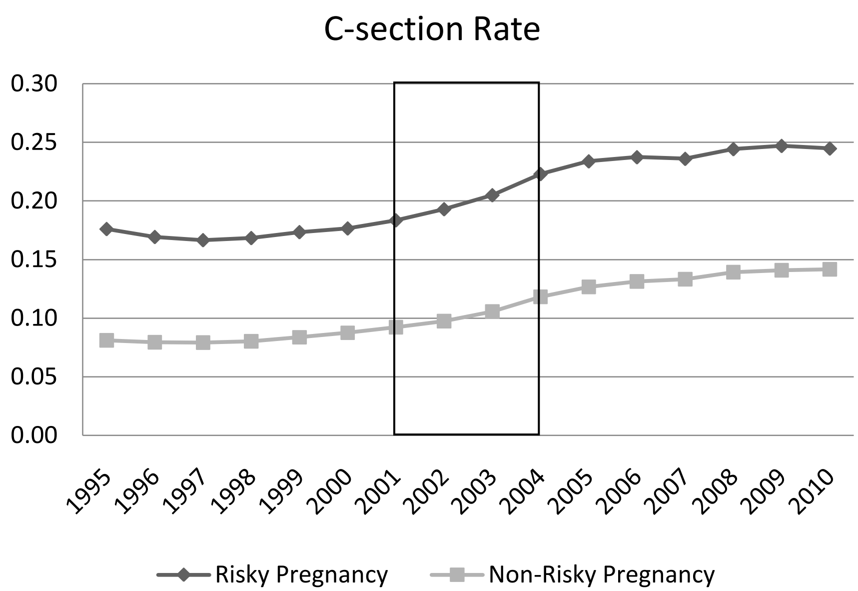
Notes: The initial Term Breech Trial (TBT) results were published in October 2000, the reversed TBT results were published in October 2004. Time 0 in the event-studies is year 2005 since the reversed TBT results were published at the end of 2004. Each dark marker is an estimate of the a^j coefficient from equation (A1) with the confidence interval on either side. Group A are term singleton breech births. Group B builds on Group A and excludes those with previous C-sections. Group C also removes first parity births, and finally, group D also removes those characterized with risky pregnancies (ie. Chronic or gestational diabetes, blood pressure, or eclampsia).

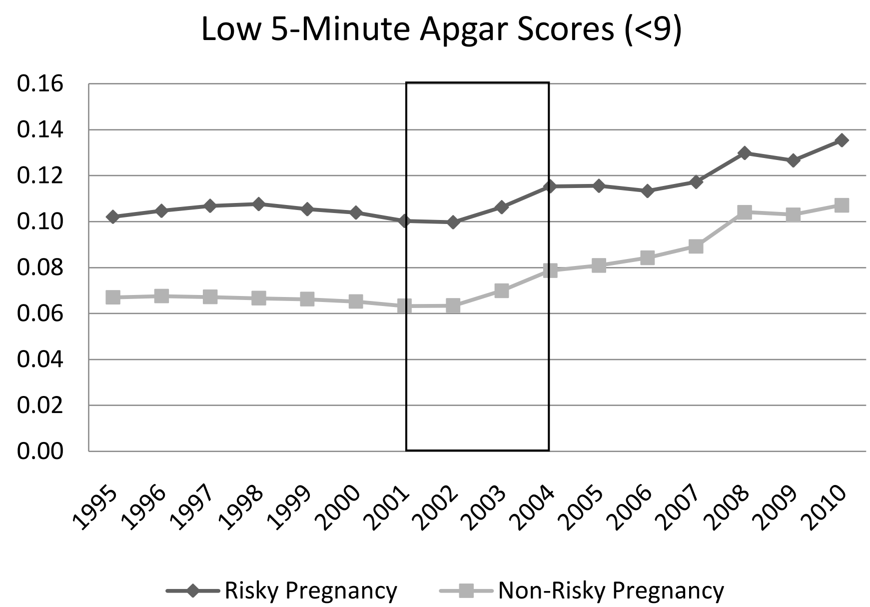
Appendix Figure B6: Change in Low Birth Weight (<2500g) - Event-Study for Breech Deliveries

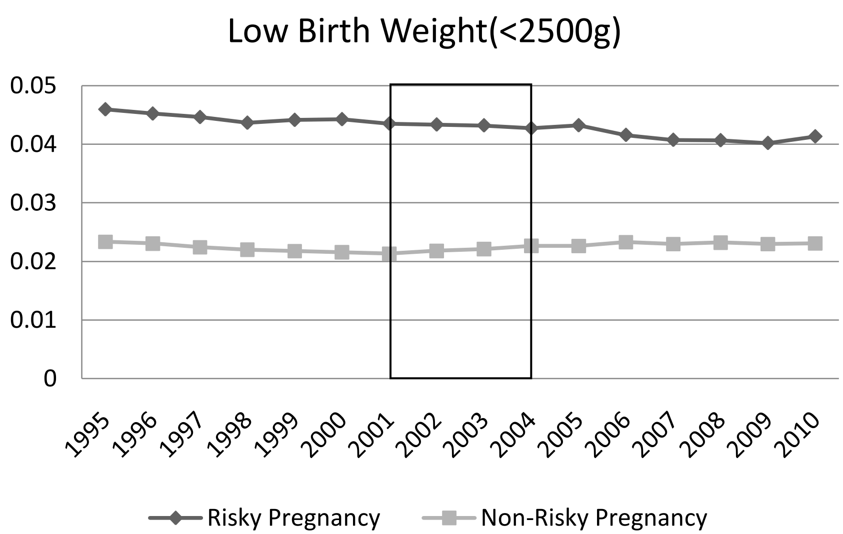


Notes: The initial Term Breech Trial (TBT) results were published in October 2000, the reversed TBT results were published in October 2004. Time 0 in the event-studies is year 2005 since the reversed TBT results were published at the end of 2004. Each dark marker is an estimate of the a^j coefficient from equation (A1) with the confidence interval on either side. Group A are term singleton breech births. Group B builds on Group A and excludes those with previous C-sections. Group C also removes first parity births, and finally, group D also removes those characterized with risky pregnancies (ie. Chronic or gestational diabetes, blood pressure, or eclampsia).

Appendix C: Falsification Test (Risky Pregnancy Trends)







Notes: These figures serve to show the parallel paths for the risky pregnancy group versus the non-risky pregnancy control group. The initial Term Breech Trial (TBT) results were published in October 2000, the reversed TBT results were published in October 2004. These figures correspond with the last column in Table 2 where the estimates for the falsification test results are displayed.

Appendix D: Outcomes from the initial and reversed TBT results on term singleton breech births (Group D)

	1	2	3	4	5
Panel A: C-section Outcome During Initial TBT Results					
Coefficient β_1	-0.004	-0.004	-0.005	-0.006	-0.006
Standard Error	0.007	0.007	0.008	0.009	0.009
P-value	0.556	0.540	0.486	0.498	0.486
Percentage Change (%)	-0.5	-0.5	-0.7	-0.8	-0.8
Observations	17436070	17436070	14011826	10044080	10044080
Panel B: C-section Outcome During Reversed TBT Results					
Coefficient β_1	-0.168	-0.166	-0.161	-0.182	-0.182
Standard Error	0.054	0.055	0.062	0.072	0.072
P-value	0.003	0.004	0.013	0.016	0.017
Percentage Change (%)	-21.4	-21.1	-20.5	-23.2	-23.2
Observations	21040479	20836251	17089810	12810002	12810002
Panel C: Low 5-Minute Apgar (<9) Outcome During Reversed TBT Results					
Coefficient β_1	0.028	0.028	0.024	0.021	0.018
Standard Error	0.009	0.009	0.009	0.010	0.010
P-value	0.004	0.005	0.014	0.054	0.077
Percentage Change (%)	25.5	25.3	21.7	19.9	18.9
Observations	18168778	18168365	16141110	11879983	11879983
Panel D: Low Birth Weight (<2500g) Outcome During Reversed TBT Results					
Coefficient β_1	-0.006	-0.006	-0.006	-0.005	-0.005
Standard Error	0.001	0.002	0.002	0.002	0.002
P-value	0.000	0.000	0.001	0.008	0.008
Percentage Change (%)	-14.1	-15.2	-13.5	-12.8	-12.9
Observations	21040162	20835934	17089810	12810002	12810002
Panel E: Included Controls					
County FE	X	X	X	X	X
Year FE	X	X	X	X	X
Mother Characteristics		X	X	X	X
Pregnancy Characteristics			X	X	X
State Characteristics				X	X
State Specific Linear Time Trends					X

Note: This table illustrates the progression of equation (1) for our preferred sample (Group D). Group D consists of non-risky term singleton births with no previous C-sections and first parity births removed. The initial Term Breech Trial (TBT) results were published in October 2000, the reversed TBT results were published in October 2004. Pre-TBT: Jan.95-Oct.00; Initial TBT: Nov.00-Oct.04; Reversed TBT: Nov.04-Dec.10. Panel A does not include the reversed TBT period, and Panel B, C, and D do not include the initial TBT period, although its inclusion does not change estimates. Each panel and column is from a separate regression. Standard errors are clustered at the state level. The percentage change is the coefficient divided by the pre-TBT (prior to October 2000) C-section rate for breech deliveries.

Appendix E: Effects of Reversed TBT on C-Sections and Infant Health Outcomes-Sensitivity Analysis

	High C-Section Rate States Removed	High C-Section Rate Counties Removed	Low C-Section Rate States Removed	Low C-Section Rate Counties Removed
Panel A: C-section Outcome During Reversed TBT Results				
Coefficient β_1	-0.194	-0.167	-0.184	-0.193
Standard Error	0.074	0.062	0.077	0.078
P-value	0.014	0.011	0.024	0.018
Percentage Change (%)	-24.7	-21.5	-23.2	-24.3
Observations	11933863	10472387	11943481	11160608
Panel B: Low 5-Minute Apgar (<9) Outcome During Reversed TBT Results				
Coefficient β_1	0.024	0.026	0.021	0.022
Standard Error	0.011	0.013	0.011	0.011
P-value	0.030	0.049	0.061	0.057
Percentage Change (%)	21.4	23.0	19.8	21.0
Observations	11006281	9794091	11016712	10263479
Panel C: Low Birth Weight (<2500g) Outcome During Reversed TBT Results				
Coefficient β_1	-0.006	-0.005	-0.005	-0.006
Standard Error	0.002	0.002	0.002	0.002
P-value	0.007	0.005	0.020	0.007
Percentage Change (%)	-13.5	-12.4	-12.0	-13.8
Observations	11933863	10472387	11943481	11160608

Note: This is for Group D which consists of non-risky term singleton births with no previous C-sections and first parity births removed. The "High C-Section Rate" is above 10 percent and consists of the District of Columbia, Louisiana, Maryland, and New York. The "Low C-Section Rate" is below 5 percent and consist of Alaska, Minnesota, Utah, and Wisconsin. The high and low C-section rate counties comprise of the top and bottom 10th percentile, which again is 10 and 5 percent, respectively. The reversed Term Breech Trial (TBT) results were published in October 2004. Pre-TBT: Jan.95-Oct.00; Initial TBT: Nov.00-Oct.04; Reversed TBT: Nov.04-Dec.10. Panel A, B, and C do not include the initial TBT period, although its inclusion does not change estimates. Each panel and column is from a separate regression. The linear probability model includes day of the week, month, year, and county fixed-effects, demographic, pregnancy, and state controls listed in Table 1. Standard errors are clustered at the state level. The percentage change is the coefficient divided by the pre-TBT (prior to October 2000) C-section rate for breech deliveries.

Appendix F: Effects of Reversed TBT on C-Sections and Infant Health Outcomes-Probit Model

	C-Section	Low 5-Minute Apgar (<9)	Low Birth Weight (<2500g)
Coefficient β_1	-0.688	0.056	-0.077
Standard Error	0.213	0.031	0.025
P-value	0.001	0.075	0.002
Marginal Effect	-0.113	0.008	-0.004
Observations	12,809,993	11,879,970	12,809,993

Note: This is for Group D which consists of non-risky term singleton births with no previous C-sections and first parity births removed. The reversed Term Breech Trial (TBT) results were published in October 2004. Pre-TBT: Jan.95-Oct.00; Initial TBT: Nov.00-Oct.04; Reversed TBT: Nov.04-Dec.10. These models do not include the initial TBT period, although its inclusion does not change estimates. Each column is from a separate regression. The probit model includes day of the week, month, year, and county fixed-effects, demographic, pregnancy, and state controls listed in Table 1. Standard errors are clustered at the state level.

Appendix G: Effects of Reversed TBT on C-Sections and Infant Health Outcomes- Annual Treatment Effects

	C-Section	Low 5-Minute Apgar (<9)	Low Birth Weight (<2500g)
Coefficient (Treated*2005)	-0.166	0.014	-0.004
Standard Error	0.093	0.010	0.002
P-value	0.083	0.156	0.079
Coefficient (Treated*2006)	-0.200	0.044	-0.004
Standard Error	0.097	0.015	0.002
P-value	0.048	0.005	0.099
Coefficient (Treated*2007)	-0.227	0.028	-0.007
Standard Error	0.073	0.016	0.002
P-value	0.004	0.084	0.001
Coefficient (Treated*2008)	-0.184	0.011	-0.006
Standard Error	0.055	0.013	0.002
P-value	0.002	0.406	0.001
Coefficient (Treated*2009)	-0.160	0.011	-0.004
Standard Error	0.065	0.010	0.002
P-value	0.018	0.283	0.007
Coefficient (Treated*2010)	-0.163	0.015	-0.005
Standard Error	0.059	0.015	0.003
P-value	0.009	0.339	0.071
Observations	12,810,002	11,879,983	12,810,002

Note: This is for Group D which consists of non-risky term singleton births with no previous C-sections and first parity births removed. The reversed Term Breech Trial (TBT) results were published in October 2004. Pre-TBT: Jan.95-Oct.00; Initial TBT: Nov.00-Oct.04; Reversed TBT: Nov.04-Dec.10. These models do not include the initial TBT period, although its inclusion does not change estimates. Each column is from a separate regression. The linear probability model includes day of the week, month, year, and county fixed-effects, demographic, pregnancy, and state controls listed in Table 1. Standard errors are clustered at the state level.

Appendix H: Discussion on Heterogeneous Results on Patient's Education and Race

We present estimates from our models by mother's education and race in Table 4. First, and not surprisingly, all the coefficients from the reversed TBT results on C-sections are statistically significant. However, some interesting results emerge. There is roughly double the decline in C-section rates for mothers with no high school diploma (33%) versus mothers with some college and higher (16%). This finding runs counter to other studies that document greater impact of medical announcements among formally educated patients. Price and Simon (2009), for example, find that new medical information on vaginal deliveries after a C-section (VBACs) had the largest impact among more educated mothers, illustrating that formal education of the patient might play a role in procedural choices. Unlike with the TBT, however, the medical study they examine was publicly announced on radio and in newspapers, and the authors argued that more educated patients may have been more likely to learn about it and to manifest certain preferences that could influence procedural choice. Jensen and Wüst (2015) also point out that the initial TBT results were quickly disseminated amongst the medical community, whereas Price and Simon's (2009) medical information was shared on public radio and in newspapers. In the case of the TBT results, both the initial findings and their eventual reversal were disseminated amongst relevant physicians, so that patient information is likely not the main factor affecting the change in C-section rates we detect.[22] We, however, note that a wider information gap between a physician and a patient may enable the physician to choose a specific delivery method more freely, such that the strength of the response measured across different

[22] Specifically, the C-section rate amongst breech deliveries dropped from 81 percent the month the reversal was published (October 2004) to 74 percent in January 2005, an almost ten percent decrease. This provides further evidence the immediate change in C-section rates is driven by physicians and not patients.

patient types may vary.[23] This could explain the larger effects we find for patients with lower educational achievement in our sample.

Table 4 also shows that Hispanic and Black mothers experienced a far larger (more than two orders of magnitude) drop in C-sections than White mothers. We subsample by race because there is evidence that physicians perceive risk of legal liability based on patient socioeconomic status (Burstin et al., 1993, Green et al., 2007, and McClellan et al., 2012), which may lead them to stay closer to guidelines with certain patients compared to others. Of course, differences in race may serve as proxies for differences in unobserved socioeconomic status, but our data do not include very rich socioeconomic information allowing us to separate the two. Interestingly, the education and racial groups that have the greater swing in C-sections do not have statistically significant effects on low Apgar scores. These results, also taken with the large confidence intervals for the event-study graphs in Appendix Figure B5, suggest that the decline in C-sections perhaps is not met with a meaningful increase in the incidence of low Apgar scores. This suggests that negative health outcomes do not necessarily follow the drop in C-sections for breech babies, at least for those groups. Perhaps unsurprisingly given the result on C-section rates for Hispanic and Black mothers, we also find a much greater reduction in the occurrence of lower birth weight among these groups. As explained previously, even among term births, the gestational period is likely to be shorter with scheduled C-sections, which could result in occurrences of low birth weight decreasing as C-section rates decrease.

[23] Johnson and Rehavi (2016) find that doctors are 10 percent less likely to perform a C-section when they are treating other doctors. This suggests that when the patient is more medically informed, C-sections are much less frequent, or at the very least that physicians do alter their procedural choice based on who they are treating and not necessarily on a purely medical diagnosis.